DIRECT METHODS IN X-RAY STRUCTURE ANALYSIS

PHASE CORRELATION AND ITS APPLICATION TO CALCULATED COSINE INVARIANTS AND DIRECT METHODS APPLIED TO DIFFERENCE STRUCTURE FACTORS

Th.E.M. van den Hark

DIRECT METHODS IN X-RAY STRUCTURE ANALYSIS

PHASE CORRELATION AND ITS APPLICATION TO CALCULATED COSINE INVARIANTS AND DIRECT METHODS APPLIED TO DIFFERENCE STRUCTURE FACTORS

Promotor: Dr. P.T. Beurskens

DIRECT METHODS IN X-RAY STRUCTURE ANALYSIS

PHASE CORRELATION AND ITS APPLICATION TO CALCULATED COSINE INVARIANTS AND DIRECT METHODS APPLIED TO DIFFERENCE STRUCTURE FACTORS

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE WISKUNDE EN NATUURWETENSCHAPPEN AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN, OP GEZAG VAN DE RECTOR MAGNIFICUS PROF. DR. A.J.H. VENDRIK, VOLGENS BESLUIT VAN HET COLLEGE VAN DECANEN IN HET OPENBAAR TE VERDEDIGEN OP DONDERDAG 17 JUNI 1976 DES NAMIDDAGS TE 4.00 UUR

DOOR

THEODOOR ERIC MARIA VAN DEN HARK

GEBOREN TE SINT HUBERT

1976

DRUK STICHTING STUDENTENPERS NIJMEGEN

Allen die op enigerlei wijze hebben bijgedragen aan de totstandkoming van dit proefschrift zeg ik hierbij van harte dank.

These investigations were supported by the 'Stichting voor Fundamenteel Onderzoek der Materie met Rontgen- en Elektronenstralen' (FOMRE) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

Aan mijn moeder
Voor Ida

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

X-ray crystallography is a powerful technique to obtain a large amount of information about molecules in the crystalline state. From nearly any compound that can be crystallized and that is stable during the collection of the X-ray data, a rather detailed picture of the crystal and molecular structure can be obtained. The accuracy and ease with which the final results are obtained are determined by several factors e.g. the quality of the X-ray data which is closely connected with the quality of the crystal from which the data are collected. In an X-ray crystal structure determination, the accuracy with which the atomic positions are determined usually ranges from a few thousandths to a few hundredths of an Ångstrom.The required accuracy varies with the purpose for which the X-ray analysis is required. If coordination problems in complexes are studied, very accurate atomic positions are not required. If, on the contrary, accurate studies of electron density distributions and atomic parameters are done, one has to do his utmost to eliminate as many factors as possible that decrease the accuracy of the results.

Before a picture of the crystal structure can be obtained from the X-ray data, the 'phase problem' (see section 1.2) has to be solved. Several techniques are available to solve this problem. Partly depending on the type of compound under investigation, a choice is made from these techniques. In this thesis so-called direct methods will be studied. Direct methods try to solve the phase problem by mainly statistical (algebraic and probabilistic) methods.

The first formulae for phase determination with direct methods were derived some 25 years ago. Some years later, the first crystal structure determinations by direct methods were reported, but it was some 15 years before direct methods were applied widely. The reason for the increase of direct methods applications around 1965 can be found in the development of practical procedures, suitable to be executed with the aid of large and fast computers, that became available at that time.Since then, the development of direct methods has been rather explosive.

Nowadays, a large fraction of all crystal structure determinations are done by direct methods. Most of them are done quite routinely. Besides this enormous quantitative increase of direct methods applications, there is a tendency to tackle crystal structures of greater complexity with direct methods, and even protein structures are subject to direct phase extension procedures.

1.2 The phase problem and direct methods

The periodic electron density in a crystal can be written as a Fourier summation:

$$\rho(\underline{r}) = V^{-1} \sum_{\underline{h}} F_{\underline{h}} \exp(-2\pi i(\underline{h}.\underline{r})) \qquad (1.1)$$

in which V represents the volume of the unit cell, $\underline{h} \equiv (h,k,l)$ is a vector whose components are the reflection indices and $\underline{r}$ is the position (in fractional coordinates) of a point in the unit cell. The complex coefficients $F_{\underline{h}} = |F_{\underline{h}}| \exp(i\phi_{\underline{h}})$ are the structure factors.

If the coefficients $F_{\underline{h}}$ were directly available from experiment, the structure of any crystal could be readily computed from eq. (1.1). Ordinarily the magnitudes $|F_{\underline{h}}|$ of the coefficients are obtainable from experiment. The phases, $\phi_{\underline{h}}$, of the structure factors, however, cannot be obtained directly from experiment; the phase information is lost in the measurement of a single reflection, but is in fact contained in the complete set of measured structure factor amplitudes. For example, it was recognized a long time ago that, if in a centrosymmetric structure two reflections with indices $\underline{h}$ and $2\underline{h}$ both had large structure factor amplitudes, the sign of the latter reflection was probably positive ($\phi_{2\underline{h}} = 0^{o}$). Harker and Kasper (1948) derived an inequality relationship which states that this probability becomes a certainty under certain conditions. The 'apparent' absence of phase information is generally refered to as the 'phase problem'. Extracting this phase information from the experimentally determined structure factor amplitudes constitutes an interesting and challenging problem for the X-ray crystallographer.

The term 'direct methods' refers to the determination of the crystal structure directly from the structure factor amplitudes without the use of chemical information, such as the presence of a heavy scatterer or

the structural knowledge of a molecular fragment. In those cases other methods (Patterson, isomorphous replacement, anomalous scattering, etc.) may provide a basis for the determination of structure from the intensities of diffraction. Commonly direct methods procedures attempt to derive the phases of the structure factors by means of statistical (algebraic and probabilistic) methods from the X-ray diffraction data. This means that the phases are obtained from relations between phases of several structure factors or from relations which link the phases directly to the structure factor amplitudes. Most formulae depend upon general principles such as the fact that the structure does consist of distinct atoms and $\rho(\underline{r})$ must have a positive value at each point in the unit cell. The relations between phases are derived supposing that the atoms are randomly distributed in the unit cell. The relations lose their significance as a function of increasing non-randomness. It is found that most initial failures to solve a crystal structure by direct methods are the result of the non-randomness of the atoms in the unit cell.

In general, direct methods are used to solve 'equal atom' structures: structures consisting of atoms of roughly the same atomic number (also frequently called 'light atom' structures). When direct methods are applied to heavy atom structures (structures that contain one or more heavy atoms in addition to a larger number of light atoms), only the positions of the heavy atoms are usually obtained. When many heavy atoms are present in the structure, it may be useful to determine the heavy atom positions by direct methods; otherwise, the heavy atom positions can, probably with less effort, be obtained from a Patterson map.

1.3 The present study

In this thesis some quite different applications of direct methods are described. In part A, a description will be given of the use of direct methods to solve equal atom structures. A procedure will be presented to solve those structures, which are hardly, if at all, solvable by existing routine direct methods techniques. The procedure is an extension of the work done on this subject by Kanters et al. (1966) and Beurskens (1965). We applied the procedure to four crystal structure determinations. Three of them will be included in part A, the fourth in part B.

In part B, procedures will be presented in which direct methods are used to speed up the determination of structures which are partially known. Originally, the procedures were developed for heavy atom structures with known heavy atom positions and unknown positions of the remaining (light) atoms. A procedure was reported by Beurskens and Noordik (1971) for the case in which the known heavy atoms are situated on special positions in centrosymmetric space groups. This procedure is now extended for heavy atoms on general positions, in both centrosymmetric and non-centrosymmetric space groups.

Recently, it was observed that these procedures are also very useful for partially known equal atom structures. When direct methods are used to solve equal atom structures (part A of this thesis), occasionally only a partial structure is obtained. Sometimes a molecule (or at least a recognisable part of it) is found at an incorrect position in the unit cell with respect to the symmetry elements of the space group. In these cases, the procedures of part B have been proved to be useful to find the remainder and/or the correct location of the structure. Examples of the application of the procedures are included in part B.

It should be emphasized that structures in non-centrosymmetric space groups, and especially in polar space groups, are in general more difficult to solve with direct methods than structures in centrosymmetric space groups. We regret that we did not have the opportunity to tackle a real problem in space group P1.

For simplicity, the compounds of which crystal structure determinations are described in this thesis will be given coded names. Often these names will be used instead of their full names.

Coded name	Chapter	Full name
'HEP.S'	5	Dinaphtho{1,2-a;1',2'-h}anthracene
'LAAR'	8	5,6,11,12-tetraphenyl-dibenzo{2-3,8-9}tricyclo{8,2,0,$0^{4,7}$}dodecadiene-2,8
'DIOX'	9	2-(di-*p*.anisylmethyl)-1,3-dioxolane
'HEPTA.NC'	11	Heptahelicene (non-centric modification)
'KRh'	13	Potassium Bis(dimethylphenylphosphine)bis-(dithiocarbonato)rhodate(III) Trihydrate

'RUBIF'	15	Bis(tetraglyme) rubidium$^+$ biphenyl$^-$
'HEPTA.C'	17	Heptahelicene (centric modification)

1.4 References

Beurskens, P.T. (1965). Thesis, University of Utrecht.

Beurskens, P.T. and Noordik, J.H. (1971). Acta Cryst. A27, 187.

Harker, D. and Kasper, J.S. (1948). Acta Cryst. 1, 70.

Kanters, J.A., Kroon, J., Beurskens, P.T. and Vliegenthart, J.A. (1966). Acta Cryst. 21, 990.

PART A

PHASE CORRELATION AND ITS APPLICATION TO CALCULATED COSINE INVARIANTS

CHAPTER 2

INTRODUCTION TO PART A OF THIS THESIS

2.1 Design of part A

In part A of this thesis, our investigations on the application of direct methods to equal atom structures are presented. If one is faced with a structural problem consisting of atoms of roughly equal atomic numbers, it is nowadays common practice to solve such a problem by direct methods. Several direct methods procedures are available which can be executed fully automatically on computers. The user of the computer program does not need to be a 'direct methods specialist', and the chemical information required to run such a program usually consists only of the chemical composition of the compound under investigation. The chance that a crystal structure of medium size complexity is solved immediately, or after only a few unsuccessful trials, by these routine methods is quite large. Nevertheless, there is a non-negligible fraction of structures that resist prolonged attempts.

When the author of this thesis started his work on direct methods in 1972, several unsuccessful attempts had been made to solve the crystal structure of 'LAAR' using several computer programs available at our laboratory at that time (Benci, 1971). The work of Kanters et al. (1966) and Hauptman et al. (1969) inspired us to use a procedure which is based on the application of a more powerful formula, to tackle this problem. We wrote the required computer programs and solved the structure. Later, the procedure was used in the structure determination of another three compounds: 'HEPTA.C', 'DIOX' and 'HEPTA.NC' , the last two crystallizing in a non-centrosymmetric space group.

The basic procedure makes use of a sigma-2 phase determination tech-

nique applied to the results of calculations with more powerful formulae (cosine invariants). Several sigma-2 procedures are available, but the use of the phase correlation procedure is prefered. For simplicity, the cosine invariant calculations are done with the old B3,0 formula (see chapter 6). In general, our procedure can easily be extended to introduce more advanced formulae, such as the MDKS-formula (Hauptman, 1972). When necessary, information from other phase relationships (such as sigma-3 and pair relationships) is also used.

In chapter 3, two frequently used sigma-2 procedures are described, along with an introduction to the phase correlation procedure, which is described in some more detail in chapter 4. Chapter 5 deals with an example of the application of the phase correlation procedure to a centrosymmetric structure ('HEP.S'). Chapter 6 is an introduction to the cosine invariant calculations. In chapter 7 the procedure using cosine invariant calculations with the B3,0 formula in phase determination with the phase correlation procedure is presented. The crystal structure determinations of 'LAAR' and 'DIOX' are presented in chapters 8 and 9,respectively. Chapter 10 deals with a description of the sigma-3 and pair relationships and the way in which the results of these relationships can be used. The crystal structure determination of 'HEPTA.NC' is given in the remaining chapter of part A.

2.2 Normalized structure factors

Many formulae in direct methods are expressed in terms of normalized structure factors, $E_{\underline{h}}$, which are theoretically defined as:

$$E_{\underline{h}} = (\epsilon_{\underline{h}}\sigma_2)^{-\frac{1}{2}} \sum_{j=1}^{N} Z_j \exp(2\pi i \underline{h}.\underline{r}_j) \qquad (2.1)$$

where N is the number of atoms in the unit cell, Z_j the atomic number of the jth atom, $\epsilon_{\underline{h}}$ a correction for symmetry enhancement of reflected intensities and

$$\sigma_n = \sum_{j=1}^{N} Z_j^n \qquad (2.2)$$

E's defined by (2.1) have to satisfy the relation $\langle |E_{\underline{h}}|^2 \rangle = 1$, when averaged over all, or an arbitrarily chosen set of vectors $\underline{h}$.

Quasi normalized structure factors, $\xi_{\underline{h}}$, are defined by:

$$\xi_{\underline{h}} = \epsilon_{\underline{h}}^{\frac{1}{2}} E_{\underline{h}} \qquad (2.3)$$

The $\xi_{\underline{h}}$'s are structure factors for a structure in which the actual atoms

are replaced by non-vibrating point atoms.

Normalized structure factor amplitudes can be calculated from the experimentally obtained $|F_{\underline{h}}|$ values by the relationship:

$$|E_{\underline{h}}| = K|F_{\underline{h}}| \left(\epsilon_{\underline{h}} \sum_{j=1}^{N} f_{j\underline{h}}^2\right)^{-\frac{1}{2}} \tag{2.4}$$

where $f_{j\underline{h}}$ is the scattering factor for the jth atom, including a correction for thermal motion and K is a scaling factor to bring the observed structure factor amplitudes to an absolute scale. The scaling factor and temperature factor are obtained by comparison of theoretically expected and experimentally observed averages of the reflected intensities as a function of the diffraction angle. (Wilson, 1942; Karle and Hauptman, 1953; Main, 1975).

For centrosymmetric and non-centrosymmetric structures the probability distributions of the normalized $|E|$'s are quite different (Wilson, 1949; Ramachandran and Srinivasan, 1959):

$$P_c(|E|)d|E| = (2/\pi)^{\frac{1}{2}} \exp(-|E|^2/2)d|E| \tag{2.5}$$

$$P_{NC}(|E|)d|E| = 2|E| \exp(-|E|^2)d|E| \tag{2.6}$$

The subscripts C and NC refer to centrosymmetric and non-centrosymmetric structures. In fig. 2.1 both functions are plotted against $|E|$. Frequently, use is made of the very dissimilar shapes of the two

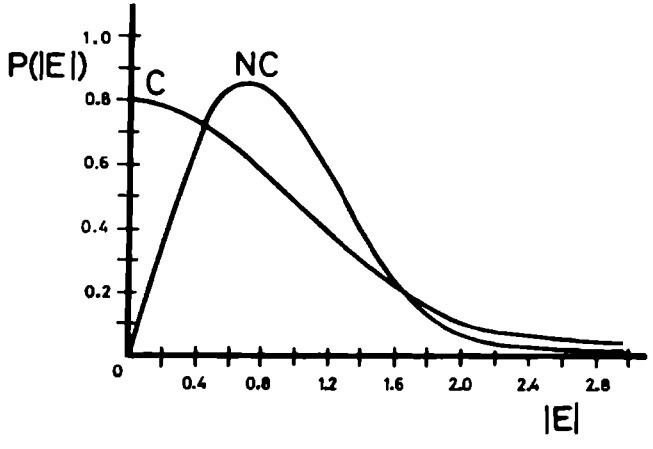


Fig. 2.1 Theoretical curves of P(|E|) against |E| for centrosymmetric (C) and non-centrosymmetric (NC) distributions.

curves to determine if a centre of symmetry is present in the structure under investigation.

2.3 The first few years of direct methods

The origin of direct methods may be traced back to the inequality relationships between structure factors, first derived by Harker and Kasper (1948). The use of the inequality relationships for phase determination is limited, however, since the reflections participating in the relationships have to satisfy very severe conditions. Such reflections are rare in crystal structures of normal size (with 100 or more atoms in the unit cell). Woolfson (1961) expected that only structures with 20 or fewer atoms in the unit cell are solvable with inequalities.

In 1952, Sayre showed that there exists an exact relationship between the structure factors of an equal atom structure. This relationship, formulated in terms of E's, reads:

$$E_{\underline{h}} = N^{\frac{1}{2}} \langle E_{\underline{k}} E_{\underline{h}-\underline{k}} \rangle_{\underline{k}} \tag{2.7}$$

Equation (2.7) has exact validity if the average is taken over an infinite or a 'very large' number of reflections. When we are interested in the phases of the E's this relationship becomes:

$$\phi(E_{\underline{h}}) = \phi\Big(\sum_{\underline{k}} E_{\underline{k}} E_{\underline{h}-\underline{k}} \Big) \tag{2.8}$$

where $\phi(E_{\underline{h}})$ denotes the phase of $E_{\underline{h}}$. It can be seen from (2.8) that, to determine any one phase, the phases and magnitudes of all other reflections need to be known. In this form eq. (2.8) is not useful for phase determination. About the same time, it was pointed out, however, that a single term of (2.8) probably indicates the correct phase of $E_{\underline{h}}$, and that the reliability of this phase indication increases with increasing magnitudes of the E's of the three reflections (Zachariasen, 1952; Cochran, 1952; Hauptman and Karle, 1953). This may be expressed as:

$$\phi(E_{\underline{h}}) \approx \phi(E_{\underline{k}}) + \phi(E_{\underline{h}-\underline{k}}) \tag{2.9}$$

where the symbol ≈ means 'probably equals'. (In those days the results were given in terms of signs, for centrosymmetric structures). Still, the phases of a number of reflections have to be known before (2.9) can be used for further phase determination. In 1953 Hauptman and Karle

derived some formulae which permitted the construction of a set of reflections with known phases to initiate application of eq.(2.9) (especially the sigma-1 and sigma-3 relationships). Several centrosymmetric crystal structure determinations, in which this approach was used, appeared in the years thereafter.

In the early sixties, the application of these formulae led to the understanding of the application of the sigma-2 formula (symbolic addition - Karle and Karle, 1963, 1966; sign correlation - Beurskens, 1964, 1965; multisolution - Woolfson, 1961). Thereafter direct methods could be applied on a larger scale.

2.4 References

Benci,P. (1971). Annual Report to FOMRE/ZWO.

Beurskens, P.T. (1964). Acta Cryst. 17, 462.

Beurskens, P.T. (1965). Thesis, University of Utrecht.

Cochran, W. (1952). Acta Cryst. 5, 65.

Harker, D. and Kasper, J.S. (1948). Acta Cryst. 1, 70.

Hauptman, H. (1972). Crystal Structure Determination. The Role of the Cosine Seminvariants. Plenum Press. New York.

Hauptman, H., Fisher, J., Hancock, H. and Norton, D.A. (1969). Acta Cryst. B25, 811.

Hauptman, H. and Karle, J. (1953). Solution of the Phase Problem, I. The Centrosymmetric Crystal, A.C.A. Monograph No. 3.

Kanters, J.A., Kroon, J., Beurskens, P.T. and Vliegenthart, J.A. (1966). Acta Cryst. 21, 990.

Karle, I.L. and Karle, J. (1963). Acta Cryst. 16, 969.

Karle, J. and Hauptman, H. (1953). Acta Cryst. 6, 473.

Karle, J. and Karle, I.L. (1966). Acta Cryst. 21, 849.

Main, P. (1975). International Summerschool on Crystallographic Computing. Prague. Lecture A2d.

Ramachandran, G.N. and Srinivasan, R. (1959). Acta Cryst. 12, 410.

Sayre, D. (1952). Acta Cryst. 5, 60.

Wilson, A.J.C. (1942). Nature, Lond. 150, 151.

Wilson, A.J.C. (1949). Acta Cryst. 2, 318.

Woolfson, M.M. (1961). Direct methods in Crystallography. Oxford University Press, New York.

Zachariasen, W.H. (1952). Acta Cryst. 5, 68.

CHAPTER 3

SIGMA-2 PROCEDURES

3.1 Formulae

The main formulae to be used in the sigma-2 procedures are listed here.

Centrosymmetric crystals

The Zachariasen-Cochran-Sayre triple product sign relationship is:

$$S(E_{\underline{h}}) \simeq S(E_{\underline{k}}) \cdot S(E_{\underline{h}-\underline{k}}) \tag{3.1}$$

where $S(E_{\underline{h}})$ means 'sign of $E_{\underline{h}}$'. A more general form of (3.1), usually refered to as the sigma-2 relationship, was given by Hauptman and Karle (1953):

$$S(E_{\underline{h}}) \simeq S(\sum_{\underline{k}} E_{\underline{k}} E_{\underline{h}-\underline{k}}) \tag{3.2}$$

The summation over $\underline{k}$ includes in general only those reflections for which both $|E_{\underline{k}}|$ and $|E_{\underline{h}-\underline{k}}|$ are large and have known signs. (A set of three reflections, $\underline{h}$, $\underline{k}$ and $\underline{h}-\underline{k}$, is usually called a triple or a sigma-2 interaction). In addition to (3.2), the probability that the sign of $E_{\underline{h}}$ is + is given by Cochran and Woolfson (1955):

$$P_+(E_{\underline{h}}) = \tfrac{1}{2} + \tfrac{1}{2}\tanh(\sigma_3\sigma_2^{-3/2}|E_{\underline{h}}| \sum_{\underline{k}} E_{\underline{k}} E_{\underline{h}-\underline{k}}) \tag{3.3}$$

Non-centrosymmetric crystals

The non-centric counterpart of (3.1) is the 'sum of angle' formula:

$$\phi_{\underline{h}} \simeq \phi_{\underline{k}} + \phi_{\underline{h}-\underline{k}} \tag{3.4}$$

in which $\phi_{\underline{h}} = \phi(E_{\underline{h}})$ represents the phase of $E_{\underline{h}}$. A more general formula (equivalent to the sigma-2 relationship (3.2)) is:

$$\phi_{\underline{h}} \simeq \phi(\sum_{\underline{k}} E_{\underline{k}} E_{\underline{h}-\underline{k}}) \tag{3.5}$$

In this formula each of the contributing triples (or sigma-2 interactions) is weighted; the weight is proportional to $|E_{\underline{h}}||E_{\underline{k}}||E_{\underline{h}-\underline{k}}|$ (frequently called the triple product). The explicite expression for the

calculation of the phase $\phi_{\underline{h}}$ is given in terms of the tangent of this phase angle; this is the so-called tangent formula:

$$\tan\phi_{\underline{h}} \approx \frac{\sum_{\underline{k}} |E_{\underline{k}}\, E_{\underline{h}-\underline{k}}| \sin(\phi_{\underline{k}} + \phi_{\underline{h}-\underline{k}})}{\sum_{\underline{k}} |E_{\underline{k}}\, E_{\underline{h}-\underline{k}}| \cos(\phi_{\underline{k}} + \phi_{\underline{h}-\underline{k}})} \qquad (3.6)$$

The summation over $\underline{k}$ in (3.5) and (3.6) includes in general only those reflections which have large $|E|$-values and known phases. An expression for the variance of $\phi_{\underline{h}}$ (a measure of the accuracy of a phase determination) is given by Karle and Karle (1966). The variance is expressed in terms of $\alpha_{\underline{h}}$, where $\alpha_{\underline{h}}$ can be calculated from:

$$\alpha_{\underline{h}}^2 = \{ \sum_{\underline{k}} A_{\underline{h}\underline{k}} \cos(\phi_{\underline{k}} + \phi_{\underline{h}-\underline{k}})\}^2 + \{ \sum_{\underline{k}} A_{\underline{h}\underline{k}} \sin(\phi_{\underline{k}} + \phi_{\underline{h}-\underline{k}})\}^2 \qquad (3.7)$$

with

$$A_{\underline{h}\underline{k}} = 2\sigma_3\sigma_2^{-3/2} |E_{\underline{h}}|\, |E_{\underline{k}}|\, |E_{\underline{h}-\underline{k}}| \approx 2N^{-\frac{1}{2}} |E_{\underline{h}}|\, |E_{\underline{k}}|\, |E_{\underline{h}-\underline{k}}| \qquad (3.8)$$

where N is the number of atoms in the unit cell, or from the equivalent expression

$$\alpha_{\underline{h}} = 2\sigma_3\sigma_2^{-3/2} |E_{\underline{h}}|\, |\sum_{\underline{k}} E_{\underline{k}}\, E_{\underline{h}-\underline{k}}| \qquad (3.9)$$

The variance of $\phi_{\underline{h}}$ is plotted as a function of $\alpha_{\underline{h}}$ in fig.3.1.

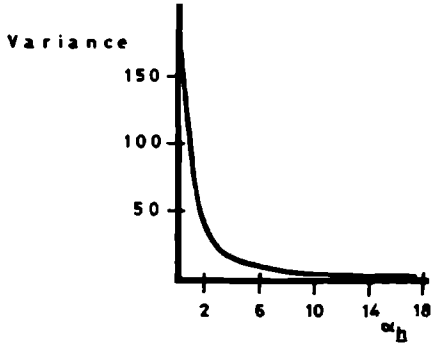


Fig. 3.1 Curve showing the variance (in squared degrees) of a phase determined from known values of other phases.

The mathematical interpretation of the variance is clear. Strong, consistent phase indications $\phi_{\underline{k}} + \phi_{\underline{h}-\underline{k}}$ are considered to be reliable

(fig. 3.2). On the other hand, weak and/or inconsistent phase indications $\phi_{\underline{k}} + \phi_{\underline{h}-\underline{k}}$ are not reliable (fig. 3.3).

Note: The expression for the variance is the non-centric generalisation of the centric probability formula (3.3).

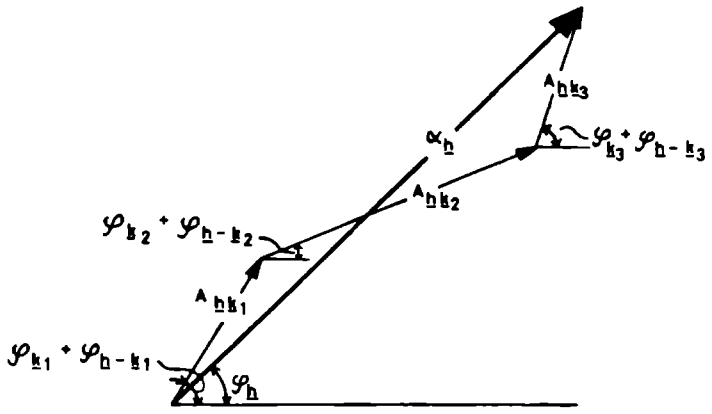


Fig. 3.2 The combination of three strong and consistent phase indications $\phi_{\underline{h}} + \phi_{\underline{h}-\underline{k}}$ *with amplitudes* $A_{\underline{h}\underline{k}}$ *to give a resultant of amplitude* $\alpha_{\underline{h}}$ *and phase* $\phi_{\underline{h}}$ *(large* $\alpha_{\underline{h}}$ *corresponds to small variance of* $\phi_{\underline{h}}$*).*

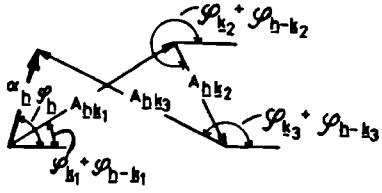


Fig. 3.3 The combination of three inconsistent phase indications $\phi_{\underline{k}} + \phi_{\underline{h}-\underline{k}}$ *with amplitudes* $A_{\underline{h}\underline{k}}$ *to give a resultant of amplitude* $\alpha_{\underline{h}}$ *and phase* $\phi_{\underline{h}}$ *(small* $\alpha_{\underline{h}}$ *corresponds to large variance of* $\phi_{\underline{h}}$*).*

In addition to the formulae (3.1) - (3.9), use is frequently made of the sigma-1 relationship (Hauptman and Karle, 1953). This relationship permits the calculation of phases (or signs) of a special group of

reflections {'cosine seminvariants' (see chapter 10)} from the absolute values of the normalized structure factors alone, using weak as well as strong reflections. The sigma-1 relationship is space group dependent. For space group $P2_1$, for example, the sigma-1 relationship reads (h01 reflections are centric):

$$S(E_{2h,0,21}) \simeq S\Big(\sum_k (-1)^k \left(|E_{h,k,1}|^2 - 1\right)\Big) \qquad (3.10)$$

with a probability that the sign of $E_{2h,0,21}$ is + :

$$P_+(E_{2h,0,21}) = \tfrac{1}{2} + \tfrac{1}{2}\tanh\Big(\sigma_3\sigma_2^{-3/2}|E_{2h,0,21}| \sum_k (-1)^k(|E_{h,k,1}|^2 - 1)\Big) \qquad (3.11)$$

3.2 The symbolic addition procedure

The symbolic addition procedure of Karle and Karle (1963, 1966) will be described here. This procedure uses letter symbols to represent the unknown phases of some reflections. To initiate the phase determination procedure, a set of reflections is chosen which comprises:
a) some reflections of which the phases are fixed (by arbitrary choices) in order to fix the origin of the electron density function and, if necessary, to select one of the enantiomorphic structures (Hauptman and Karle, 1953, 1956).
b) reflections, if any, of which the phases are 'known' for one reason or another (sigma-1 relationship, packing considerations, etc.).

The phase determination is then started using the relationships given in section 3.1. Phases that are found with a high reliability are accepted and used for further phase determination. In a step by step fashion, additional unknown phases, now expressed in terms of letter symbols, are added to the set of initial choices. The phase determination is continued until the majority of reflections with larger $|E|$-values has been phased (i.e. absolute phases or phases expressed in terms of letter symbols). The number of letter symbols is kept as small as possible. Exceptionally, relations between the letter symbols are used to eliminate one or more symbols. At the end a multiple solution is usually obtained: for p unknown symbols, at most n^p different solutions need to be considered (for centric structures n=2; for non-centric structures n is usually not more than 4 (see section 3.3)).

A problem in the symbolic addition procedure is the acceptance of a

phase for a reflection when different symbols are given for the same reflection. Moreover, single phase indications usually have to be trusted at the beginning of the phase determination process; if one or more steps are 'wrong' at this stage, the correct solution to the phase problem may easily be missed.

3.3 The multiple solution procedure

The most popular procedure for routine structure determination is certainly the multiple solution procedure, as it is introduced in the computer program MULTAN (Germain et al., 1968, 1970, 1971). Much attention has been paid in the MULTAN system to the selection of a good starting set: a starting set from which phases can be determined with maximum reliability. Such a starting set must lead as fast as possible to multiple phase indications, while, if necessary, the first few phases are determined with the 'strongest' single indications. Selection of the starting set is done with the CONVERGENCE map.

The starting set consists of:

a) reflections to fix the origin (and enantiomorph)
b) reflections of which the phases are known for one reason or the other (sigma-1, packing, etc.)
c) some other reflections

The phases of the reflections from category c) are unknown. Instead of their phases being represented by symbols, they are given explicit numerical values. For example, a reflection of which the phase is restricted to the values 0° or 180° is given each of these two values in turn; a reflection with an unrestricted phase is usually given each of the values 45,135,225 and 315° in turn. For each of the combinations of phases for the reflections from category c), the phase determination process is fully executed, and criteria are calculated to indicate the most probable order in which the different solutions should be investigated.

A possible reason for failure of the procedure as described here is that several single phase indications usually have to be trusted in the initial steps of the phase determination. In addition, the phase determination process has to be repeated in full for all of the possible solutions, which, in principle, is a waste of computer time. (In practice, however, MULTAN produces good and fast results.)

3.4 The phase correlation procedure (introduction)

The sign correlation procedure was developed in the early sixties (Beurskens, 1964). At that time, the procedure was applicable only to centrosymmetric structures. (Henceforth the name sign correlation procedure will be used only when explicitly refering to centrosymmetric structures; otherwise, the more general name phase correlation procedure will be used.) The centric procedure was programmed for the IBM 1620 computer (Beurskens, 1963) and successfully applied to several centric structures (Sax et al., 1965). Thereafter, an automatic program in ALGOL-60 was written (Beurskens, 1965). In this chapter we will introduce the phase correlation procedure as a non-centric generalisation of the sign correlation procedure.

While in the symbolic addition and multiple solution procedures single phase indications often have to be trusted in the initial stage of the phase determination process, in the phase correlation procedure all emphasis is laid on avoiding single phase indications. As a consequence, it is necessary to bring into the calculations as many reflections as possible right from the beginning of the procedure. Usually a starting set of 20 or more reflections is chosen, of which the phases are represented by letter symbols. During the phase determination procedure relations between the letter symbols are found and used to eliminate symbols. Thus, while increasing the number of reflections that take part in the analysis, the number of unknown symbols is gradually reduced. A detailed description of the procedure is given in the following chapter.

3.5 References

Beurskens, P.T. (1963). Technical Report on Sign Correlation by the Sayre Equation. Crystallography Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania.

Beurskens, P.T. (1964). Acta Cryst. 17, 462.

Beurskens, P.T. (1965). Thesis, University of Utrecht.

Cochran, W. and Woolfson, M.M. (1955). Acta Cryst. 8, 1.

Germain, G., Main,P. and Woolfson, M.M. (1970). Acta Cryst. B26, 274.

Germain, G., Main, P. and Woolfson, M.M. (1971). Acta Cryst. A27, 368.

Germain, G. and Woolfson, M.M. (1968). Acta Cryst. B24, 91.

Hauptman, H. and Karle, J. (1953). Solution of the Phase Problem, I.

The Centrosymmetric Crystal, A.C.A. Monograph No. 3.
Hauptman, H. and Karle, J. (1956). Acta Cryst. 9, 45.
Karle, I.L. and Karle, J. (1963). Acta Cryst. 16, 969.
Karle, J. and Karle, I.L. (1966). Acta Cryst. 21, 849.
Sax, M., Beurskens, P.T. and Chu, S. (1965). Acta Cryst. 18, 252.

CHAPTER 4

THE PHASE CORRELATION PROCEDURE

4.1 Introduction

In the past few years, new insights have been obtained into several essential points of the sign correlation procedure. A new automatic computer program for the execution of the sign correlation procedure has been written in FORTRAN IV, and the applicability of the phase correlation procedure to phase determination for non-centric structures has been investigated. The procedure has proved to be a powerful one, particularly in combination with cosine invariant calculations (to be discussed in chapter 7).

4.2 How it works

Definitions

- H is the set of reflections $\underline{h}$ of which the phase is known, either absolutely or expressed in terms of one or more letter symbols (The set includes all symmetry related reflections).
- Hi ≡ H1, H2, H3, ... reflections that have been assigned a letter symbol Pi (≡ P1, P2, P3, ...) to represent the phase. These 'initial choices' form a subset of the set H.
- The symbol Si (≡ S1,S2, S3, ...) denotes the sign of a reflection Hi in the sign correlation procedure.
- H + H' denotes any possible vector sum $\underline{h} + \underline{h}'$, using also symmetry related reflections.
- K is a selected set of reflections $\underline{k} = \underline{h} + \underline{h}'$ that are to be used as 'temporarily accepted' reflections (certainly not all possible reflections $\underline{h} + \underline{h}'$).
- L is the set of all reflections $\underline{l} = \underline{h} + \underline{k}$ or $\underline{k} + \underline{k}'$, and all reflections $\underline{h} + \underline{h}'$ (thus K is a subset of L).

Selection of the starting set

Some 10 to 30 reflections Hi are chosen, with $|E|$-values as large as possible, such that there are no sigma-2 interactions {= phase relationships (3.1) or (3.4) } between these reflections. Origin (and enantiomorph) determining reflections may also be added to the starting set.

Frequently, however, the origin (and enantiomorph) will be fixed in a later stage of the phase determination: some letter symbols are eliminated by giving them an arbitrarily chosen phase. Other known phases, such as sigma-1 results etc., may also be added.

Phase generation

All possible pairs of the reflections $\underline{h}$ are considered, to obtain (symbolic) phases for the reflections $\underline{h} + \underline{h}'$. The reflection set K is selected from reflections with large $|E|$-values of which the phase has been found with a large reliability {large $\alpha_{\underline{h}}$, see eq, (3.9)}. In case different symbols are found for the same reflection, the most reliable phase indication is chosen. The phases for these reflections are 'temporarily accepted'. The phase generation is continued to obtain phases for the reflections $\underline{h} + \underline{k}$ and $\underline{k} + \underline{k}'$. The resulting reflection set L is now analyzed in order to find new acceptable phases and relations among the letter symbols (correlation equations). Some reflections of L may have been found from many independent phase indications with the same phase. This means that even in case some of the temporarily accepted phases are incorrect the phases of these reflections of the set L can hardly be incorrect. These reflections are now 'definitively accepted' and added to the set H. The majority of the newly accepted reflections were also present in the set K. They are now found with a much better reliability; some reflections of the set K are not accepted. Some altogether new reflections (i.e. reflections which were not accepted temporarily) may also be accepted. A new cycle of phase correlation with the expanded set H may now be started (i.e. recalculation of the sets K and L).

Correlation equations

In each cycle of the phase correlation procedure the sets of reflections H and L are screened to find relations between the symbols; the phase of many reflections will be expressed by more than one symbol giving such relations (see for examples chapter 5). A relation which has been found with a high reliability, and very frequently with no contradictions, is selected and may be accepted to be true. This result is called a correlation equation. The correlation equation is used to elimi-

nate one of the unknown symbols Pi. If more correlation equations are acceptable, more symbols can be eliminated.

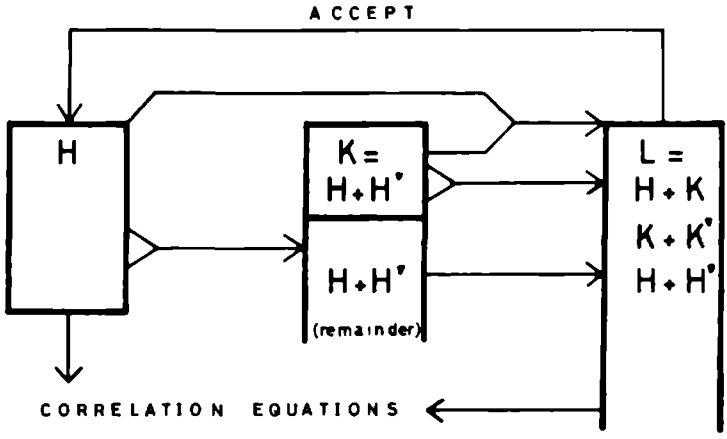


Fig. 4.1 Diagram showing the cycle phase correlation procedure.

End of the phase correlation procedure

The iterative procedure of expanding the set H and eliminating symbols Pi in an alternative sequence, usually leads to a point where no more correlation equations or reliable phases from the set L can be accepted. Sometimes, some symbols are used to express the phases of only a few reflections; these are rejected (rejected symbols are withdrawn from the procedure without expressing them in terms of other symbols) or eliminated using less reliable correlation equations. When the origin is not completely fixed, one or more symbols may be eliminated to complete the origin fixation. In the case that some symbols remain unknown, a multiple solution to the phase problem is obtained. Sometimes, it makes sense to conclude the phase generation for each of the solutions separately; all phases expressed in terms of letter symbols are replaced by numerical values before applying the tangent formula.

4.3 Quadruples and quartets in relation to the phase correlation procedure

In the phase correlation procedure large numbers of 'quadruples' and 'quartets' are used implicitly.

A 'quadruple' (de Vries (1965) used the name correlation equation) is a relationship between four triples, which can be generated from six different reflections (de Vries, 1963, 1965 ; Viterbo and Woolfson, 1973). Table 4.1 gives the simple form of a quadruple. Because of Friedel's law all ϕ's cancel out in pairs and the identity $\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 = 0^\circ \pmod{360^\circ}$ is satisfied (for centric structures: $T_1 \cdot T_2 \cdot T_3 \cdot T_4 = +$). This means that if the sum of the phases of the reflections in one of the four triples deviates much from the expectation value of 0°, this deviation must be compensated by large deviations in the other triples. For centric structures, only an even number of triples in a quadruple can be incorrect.

Table 4.1 The simple form of a quadruple.

In terms of phases

$$\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{-\underline{h}-\underline{k}} = \Phi_1$$

$$\phi_{\underline{h}+\underline{k}} + \phi_{\underline{1}} + \phi_{-\underline{h}-\underline{k}-\underline{1}} = \Phi_2$$

$$\phi_{-\underline{k}} + \phi_{-\underline{1}} + \phi_{\underline{k}+\underline{1}} = \Phi_3$$

$$\phi_{-\underline{h}} + \phi_{\underline{h}+\underline{k}+\underline{1}} + \phi_{-\underline{k}-\underline{1}} = \Phi_4$$

$$\Phi_1, \Phi_2, \Phi_3, \Phi_4 \approx 0^\circ$$

$$\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 = 0^\circ \pmod{360^\circ}$$

In terms of signs

$$S_{\underline{h}} \cdot S_{\underline{k}} \cdot S_{-\underline{h}-\underline{k}} = T_1$$

$$S_{\underline{h}+\underline{k}} \cdot S_{\underline{1}} \cdot S_{-\underline{h}-\underline{k}-\underline{1}} = T_2$$

$$S_{-\underline{k}} \cdot S_{-\underline{1}} \cdot S_{\underline{k}+\underline{1}} = T_3$$

$$S_{-\underline{h}} \cdot S_{\underline{h}+\underline{k}+\underline{1}} \cdot S_{-\underline{k}-\underline{1}} = T_4$$

$$T_1, T_2, T_3, T_4 \approx +$$

$$T_1 \cdot T_2 \cdot T_3 \cdot T_4 = +$$

Quadruples occur in the phase correlation procedure if, for example, two reflections of set H and n couples of two temporarily accepted reflections lead to the same (symbolic) phase for a reflection of set L. In this case, 3n+1 triples have been used of which n quadruples can be con-

structed, all with one and the same triple in common.

A 'quartet' is a relationship between the phases of four reflections of which the sum of the indices equals zero { sigma-5 relationship - Hauptman and Karle, 1953; coincidence of the second kind - de Vries, (1965)}. In case the |E|-values of the four participating reflections and of the three 'cross-terms' are large, the sum of the four phases probably equals a value near 0° rather than 180° {strengthened quartet - Schenk, (1973) }.

Table 4.2 The general form of a strengthened quartet.

In terms of phases	In terms of signs
$\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{-\underline{h}-\underline{k}-\underline{l}} \simeq 0^{\circ}$	$S_{\underline{h}} \cdot S_{\underline{k}} \cdot S_{\underline{l}} \cdot S_{-\underline{h}-\underline{k}-\underline{l}} \simeq +$
if $\lvert E_{\underline{h}} \rvert$, $\lvert E_{\underline{k}} \rvert$, $\lvert E_{\underline{l}} \rvert$, $\lvert E_{\underline{h}+\underline{k}+\underline{l}} \rvert$,	
$\lvert E_{\underline{h}+\underline{k}} \rvert$, $\lvert E_{\underline{k}+\underline{l}} \rvert$, $\lvert E_{\underline{h}+\underline{l}} \rvert$ all are large	

Recently, it was observed that if the |E|-values of the three cross-terms are very small, the sum of the four phases tends to be 180° instead of 0° { negative quartets - Hauptman, 1974a, 1974b, 1975; Schenk, 1974 }.

The quartets occuring in the phase correlation procedure have two cross-terms with large |E|-values; the third cross-term is not checked. Examples of quadruples and quartets are given in chapter 5.

4.4 The computer program SCOR (sign correlation)

The computer program SCOR, written in FORTRAN IV, is able to perform the sign correlation procedure for centrosymmetric triclinic, monoclinic and orthorhombic space groups (except Fddd). SCOR can carry out the sign determination in a complete automatic mode, although intervention by the user of the program is possible. A number of parameters controling some key steps in the program (selection of reflections to be accepted temporarily or definitively and selection of correlation equations to be accepted) have been given default values, which may be changed by the user of the program. In each cycle of sign correlation many thousands of triples need to be handled. These are generated every cycle again by a fast triple generation procedure. In the computer program the letter symbols

are represented by binary codes in a similar way as described by Beurskens (1965). The computer time required by SCOR is comparable with the time required by the program MULTAN when the sign determination is done for a reasonable number of solutions (say 32). For more difficult problems, where the sign determination has to be done very carefully, the required computer time will increase.

At present we are developing a system of automatic direct methods computer programs which is called the NIJDIR system (see Appendix A_1). SCOR will be one of the programs of that system.

4.5 References

Beurskens, P.T. (1965). Thesis, University of Utrecht.

Hauptman, H. (1974a). Acta Cryst. A30, 472.

Hauptman, H. (1974b). Acta Cryst. A30, 822.

Hauptman, H. (1975). Acta Cryst. A31, 680.

Hauptman, H. and Karle, J. (1953). Solution of the Phase Problem, I. The Centrosymmetric Crystal, A.C.A. Monograph No.3.

Schenk, H. (1973). Acta Cryst. A29, 77.

Schenk, H. (1974). Acta Cryst. A30, 477.

Viterbo, D. and Woolfson, M.M. (1973). Acta Cryst. A29, 205.

de Vries, A. (1965). Acta Cryst. 18, 473.

de Vries, A. (1963). Thesis, University of Utrecht.

CHAPTER 5

'HEP.S'

5.1 Crystal structure of 'HEP.S'

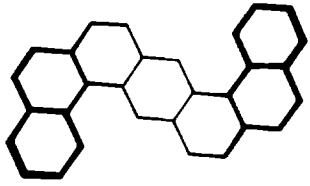
The paper published in Cryst. Struct. Comm. (1975) 4, 281 - 284 will be reproduced in this section.

DINAPHTHO{1,2-a;1',2'-h}ANTHRACENE, $C_{30}H_{18}$

B.G.M.C. Hummelink-Peters, Th.E.M. van den Hark,
J.H. Noordik and Paul T. Beurskens

Crystallography Laboratory, University of Nijmegen,
Toernooiveld, Nijmegen, The Netherlands

Preliminary information. This structure was solved by accident. We intended to solve the crystal structure of heptahelicene, prepared by Laarhoven, 1970. Recrystallization gave only one crystal, which was used for the X-ray measurements; it was then found that the impurity was crystallized. The molecule is far from planar because of intramolecular hindrance.

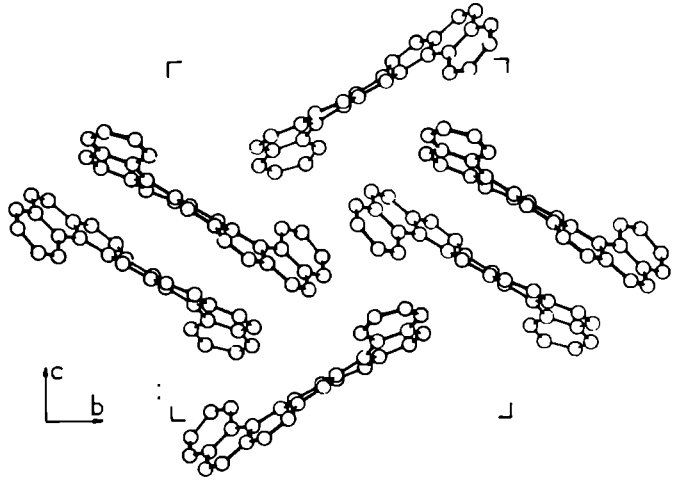


Crystal data. M.W. = 378.47. From single crystal diffractometry, λCuKα = 1.5418Å: Monoclinic, a = 8.167(3)Å, b = 15.000(5)Å, c = 15.641(5)Å, β = 90.43(2)°. Space group $P2_1/c$, Z = 4, D_x = 1.31g/cm^3.

Intensity data, structure determination and refinement. Intensity data of 2409 symmetry independent reflections up to $\sin\theta/\lambda = 0.53$Å^{-1} were measured on an automatic NONIUS diffractometer (θ-2θ scan). 1723 reflections had $I>3\sigma_c(I)$; ($\sigma_c(I)$ based on counting statistics) and were used in the refinement. The structure was solved by direct methods using the program MULTAN of Germain, Main and Woolfson (1971). The most probable solution (out of 8) was correct. The atomic positional and anisotropic thermal parameters of the carbon atoms were refined by full-matrix least-squares

techniques, minimizing the function $\Sigma\ w\ \{|F_o| - k|F_c|\}^2$, with $w = \{\sigma_c(F_o)^2 + (0.05\ F_o)^2\}^{-1}$. At R = 0.046 the hydrogen atoms with the exception of H_1, H_9, H_{10} and H_{18} were introduced at calculated positions with a C-H bond length of 1.084Å. From a difference Fourier map calculated from 392 reflections with $\sin\theta/\lambda < 0.30$Å^{-1}, the positions of H_1, H_9, H_{10} and H_{18} were found. (The numbering of the hydrogen atoms refers to their parent carbon atoms). All hydrogen atoms were included in the structure factor calculations with a temperature factor of B = 5.0Å^2, but they were not refined. After two more refinement cycles the final conventional R-factor is 0.044 for 1723 non-zero reflections.

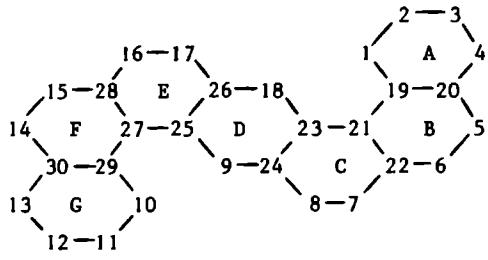
Comments. The molecule is of point symmetry 2 (C_2); the two-fold axis, however, is not a crystallographic two-fold axis. The root-mean-squares deviation from true two-fold symmetry (refered to the bond distances) is equal to the estimated standard deviation (calculated from the positional parameters). We therefore give the averaged bond distances and angles. As we found earlier in hexahelicenes and tribenzoheptahelicene (Van den Hark et al., 1974, and ref. therein) and in benzo[c]phenanthrene (Hirschfeld et al., 1963) the C-C bonds in the inner 'core' are lengthened and the peripheral bonds are shortened. The interplanar angles are A,B = 9.92°, B,C = 9.04°, C,D = 12.33°, D,E = 11.46°, E,F = 7.64° and F,G = 8.73°. Torsion angles in C(1)-C(19)-C(21)-C(23)-C(18) are 18.5 and 22.2°, in C(10)-C(29)-C(27)-C(25)-C(9) 18.7 and 18.0°. This compares very well with the results for benzo[c]phenanthrene. The distance C_1-C_{18} is 3.048(3)Å

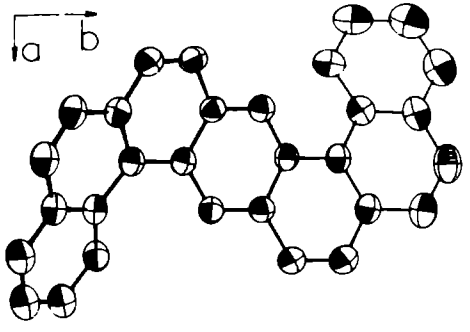


and C_9-C_{10} is 3.001(3)Å. The distance H_1-H_{18} is 2.18Å and H_9-H_{10} is 2.19Å. Comparable distances in benzo[c]phenanthrene are C-C' = 3.030Å and H-H' = 2.04Å.

Averaged bond distances and angles, C-C-C bond angles in the range 117.5 - 122.5° are not given; esd's: C-C = 0.003Å, C-C-C = 0.2°.

C-C	Å
18-23	1.398
23-21	1.458
21-19	1.451
19- 1	1.420
1- 2	1.380
2- 3	1.406
3- 4	1.362
4-20	1.408
20- 5	1.420
5- 6	1.346
6-22	1.431
22- 7	1.435
7- 8	1.350
8-24	1.434
24- 9	1.396
19-20	1.423
21-22	1.400
23-24	1.434





C-C-C	degrees
1-19-21	123.7
19-21-23	124.1
21-23-18	124.1
23-18-26	122.6
18-23-24	116.3

H-C	Å
1- 1	1.14
18-18	1.08

H-C-C	degrees
1- 1- 2	116
1- 1-19	122
18-18-23	120
18-18-26	117

Atomic coordinates. Two atoms given on one row, are related by a local two-fold axis. Esd: estimated standard deviation for carbon atoms. Hydrogen atoms on calculated positions are omitted.

Atom	x	y	z	Atom	x	y	z
(esd:)	0.0004	0.0002	0.0002	(esd:)	0.0004	0.0002	0.0002
C(1)	0.0528	0.0590	0.2568	C(10)	0.5896	-0.3279	0.4650
2	-.0730	.0998	.2116	11	.7170	-.3866	.4765
3	-.0619	.1904	.1893	12	.7172	-.4470	.5452
4	.0773	.2369	.2082	13	.5839	-.4511	.5969
5	.3600	.2430	.2607	14	.3040	-.4056	.6349
6	.4935	.2010	.2920	15	.1699	-.3564	.6192
7	.6288	.0699	.3547	16	.0297	-.2304	.5477
8	.6256	-.0138	.3863	17	.0291	-.1592	.4949
18	.1790	-.0524	.4064	9	.4700	-.1402	.4424
19	.1948	.1075	.2804	29	.4538	-.3266	.5214
20	.2096	.1969	.2512	30	.4485	-.3936	.5858
21	.3319	.0677	.3263	27	.3180	-.2647	.5148
22	.4820	.1124	.3245	28	.1741	-.2837	.5595
23	.3236	-.0155	.3741	25	.3247	-.1814	.4671
24	.4728	-.0584	.3992	26	.1768	-.1310	.4537
H(1)	.0373	-.0162	.2652	H(10)	.5829	-.2909	.4019
18	.0617	-.0111	.4038	9	.5801	-.1614	.4665

Acknowledgements. The autors thank Mr. J.M.M. Smits for technical assistance. One of us (v.d. Hark) acknowledges support of the Dutch Foundation for Pure Research, ZWO/FOMRE.

References.

Germain, G., Main, P. and Woolfson, M.M. (1971). Acta Cryst. A27,368.

Van den Hark, Th.E.M., Noordik, J.H. and Beurskens, P.T. (1974). Cryst. Struct. Comm. 3, 443.

Hirshfeld, F.L., Sandler, S. and Schmidt, G.M.J. (1963). J. Chem. Soc. 2108.

Laarhoven, W.H., Cuppen, Th.J.H.M. and Nivard, R.J.F. (1970). Tetrahedron 26, 4865.

Received: 2 January 1975

5.2 Application of the sign correlation procedure to 'HEP.S'

At the time we solved the structure of 'HEP.S' routinely by the program MULTAN - see section 5.1 -, we had just finished the program SCOR, and we decided to check the performance of the new program on the data of 'HEP.S'. In this section, we report some numerical results in order to illustrate the sign correlation procedure and also to show the relationship of the procedure with quadruples and quartets. SCOR was applied to all 388 reflections with $|E| > 1.3$. Thus the triples used in the procedure have a triple product of at least 2.2, or a probability that they satisfy the sigma-2 relationship of at least 59.7%. Symmetry relationship for space group $P2_1/c$: $S_{h\bar{k}l} = (-1)^{k+l} S_{hkl}$.

The starting set

A starting set of 26 reflections Hi was chosen from the strongest $|E|$-values, such that there exist no triples Hi + Hj = Hk; these initial reflections were given symbols Si. During the selection of this starting set, all reflections (Hi + Hj) are calculated. The algorithm used is explained by its results, as shown in table 5.1.

Sign generation and search for correlation equations

From all reflections (Hi + Hj) (that is, from the signs calculated from the starting set) the most probable new signs are selected and denoted K1, K2, K3, ... (see table 5.1). Preset control parameters led to the selection of 134 reflections k, which were used to calculate symbolic signs for more reflections. Inspection of the results for the reflections of set L showed that some correlation equations occured very frequently with few or no contradictions, and with high probabilities. Six correlation equations were accepted and used to eliminate six symbols. To continue the sign generation, a symbolic sign of a reflection of set L was definitively accepted when it was found by at least 10 consistent triples with sufficiently high probabilities. In this way, 18 reflections were accepted and added to the set of initial choices (set H).

The results for one reflection of set L are given in table 5.2. The reflection 7 1 $\bar{1}$ has been found 8 times with the symbolic sign S1.S2: once directly from starting set reflections, and seven times from temporarily accepted reflections. To indicate the same symbolic sign for the reflection 7 1 $\bar{1}$ eight times, 22 triples have been used.

Table 5.1 Signs for the strongest 20 reflections of 'HEP.S'. Among the 20 reflections are 16 starting set reflections, denoted H1, H2, ..., H16, and 4 reflections from set K, denoted K1, K2, K3, K4. The symbolic signs for the 4 reflections Ki are given in order of decreasing probability.

	reflection			symbolic sign(s)
4	8	$\bar{6}$	H1	S1
3	9	5	H2	S2
3	12	1	H3	S3
7	1	$\bar{1}$	K1	S1.S2 , -S3.S4 , S6.S14 , S10.S13
4	11	$\bar{2}$	H4	S4
2	12	5	H5	S5
4	8	8	H6	S6
3	8	8	H7	S7
4	8	$\bar{7}$	H8	S8
1	8	12	H9	S9
0	2	3	H10	S10
7	0	2	K2	S1.S7
2	7	10	H11	S11
0	2	4	H12	S12
7	1	2	H13	S13
3	7	9	K3	S2.S12 , S8.S13
3	7	$\bar{9}$	H14	S14
3	11	2	K4	-S2.S10
4	13	0	H15	S15
5	7	$\bar{6}$	H16	S16

If the sign of the reflection 7 1 $\bar{1}$ is incorrectly indicated by the symbolic sign S1.S2, at least 8 of the 22 triples must be incorrect. This can hardly be true, and the symbolic sign S1.S2 for the reflection 7 1 $\bar{1}$ may be trusted. Nevertheless, our preset limit was 10 consistent sigma-2 indications. After accepting the six correlation equations, however, the sign S1.S2 for the reflection 7 1 $\bar{1}$ was found 20 times consistently and was definitively accepted.

Table 5.2 Sign correlation results for the reflection 7 1 $\bar{1}$.

a. Reflections used to give eight times the sign S1.S2.

reflections Ki			sign		result			sign
7	1	$\bar{1}$	S1.S2	→	7	1	$\bar{1}$	S1.S2
3	11	2	-S2.S10	→	7	1	$\bar{1}$	S1.S2
4	10	$\bar{3}$	S1.S10					
2	1	$\bar{7}$	S2.S9	→	7	1	$\bar{1}$	S1.S2
5	0	6	S1.S9					
7	2	3	S2.S4	→	7	1	$\bar{1}$	S1.S2
0	3	4	S1.S4					
1	2	$\bar{5}$	S2.S11	→	7	1	$\bar{1}$	S1.S2
6	1	4	-S1.S11					
3	7	$\bar{9}$	S2.S12	→	7	1	$\bar{1}$	S1.S2
4	6	$\bar{10}$	S1.S12					
6	3	$\bar{5}$	S2.S19	→	7	1	$\bar{1}$	S1.S2
1	2	4	S1.S19					
5	3	$\bar{10}$	S2.S5	→	7	1	$\bar{1}$	S1.S2
2	4	$\bar{11}$	-S1.S5					

b. Other symbolic signs for reflection 7 1 $\bar{1}$.

-S3.S4	6 indications
-S8.S17	4 indications
S6.S14	3 indications
S10.S13	3 indications
S7. S23	2 indications

plus 7 symbolic signs indicated only once

c. From the symbolic signs in tables 5.2a and 5.2b follow the symbolic relations (in order of decreasing probability):

S1.S2 = -S3.S4

S1.S2 = -S8.S17

S1.S2 = S6.S14

S1.S2 = S10.S13

S3.S4 = S8.S17

etc.

Second cycle

The basic set for the second cycle is formed by the 26 initial reflections and the 18 newly accepted reflections (e.g. the reflection 7 1 $\bar{1}$, formerly denoted K1, will now be denoted H27). Further intermediate results are given in table 5.3

Table 5.3 Intermediate results of the sign correlation procedure as applied to 'HEP.S'.

cycle no.	no. of refl. H	no. of refl. K	eliminated symbols	no. of definitively accepted refl.
1	26	134	S4,S8,S12,S13, S14,S19	18
2	44	196	S9,S11,S15,S16, S17,S18,S20, S21,S24,S26	58
3	99	146	S6,S7	55
4	154	180	-	-

The end of the sign correlation procedure

The origin was fixed by giving to the symbols S2, S3 and S10 a plus sign. Three symbols, S22, S23 and S25, were rejected because only a small number of reflections had symbolic signs dependent on these symbols. The sign correlation procedure was stopped when the set H contained 154 reflections; for the large majority of the remaining reflections probable symbols (or absolute signs) were calculated. Two symbols S1 and S5 were the remaining unknown symbols. The correlation equations found at the end of the procedure were not reliable enough to be accepted, so a fourfold solution to the phase problem was obtained; one of these, of course, was correct.

Quadruples

The generation of symbolic signs, right from the initiation of the procedure, is directly related to the existence of quadruples among the triples that are used. As an example, the reflection 7 1 $\bar{1}$, temporarily used as K1, was found seven times with the same symbolic sign, and (together with the 3 times 7 triples involved) it is possible to construct 7 quadruples. Two quadruples are shown in table 5.4 (see also section 4.3). Many of the reflections used in the initial stage of the sign cor-

relation procedure are strong, giving rise to strong and reliable quadruples.

Table 5.4 Some quadruples (only starting set reflections from table 5.1 and temporarily accepted reflections from table 5.2 are used).

QI:	$S_{4\ \bar{8}\ \bar{6}}$	$\cdot S_{3\ 9\ 5}$	$\cdot S_{\bar{7}\ \bar{1}\ 1}$	$\simeq$ (S1).(S2).(S1.S2)
	$S_{7\ 1\ \bar{1}}$	$\cdot S_{\bar{3}\ \bar{11}\ \bar{2}}$	$\cdot S_{\bar{4}\ 10\ 3}$	$\simeq$ (S1.S2).(-S2.S10).(-S1.S10)
	$S_{\bar{3}\ \bar{9}\ \bar{5}}$	$\cdot S_{0\ \bar{2}\ 3}$	$\cdot S_{3\ 11\ 2}$	$\simeq$ (S2).(-S10).(-S2.S10)
	$S_{\bar{4}\ 8\ 6}$	$\cdot S_{0\ 2\ \bar{3}}$	$\cdot S_{4\ \bar{10}\ \bar{3}}$	$\simeq$ (S1).(-S10).(-S1.S10)
QII:	$S_{4\ \bar{8}\ \bar{6}}$	$\cdot S_{3\ 9\ 5}$	$\cdot S_{\bar{7}\ \bar{1}\ 1}$	$\simeq$ (S1).(S2).(S1.S2)
	$S_{7\ 1\ \bar{1}}$	$\cdot S_{\bar{2}\ \bar{1}\ 7}$	$\cdot S_{\bar{5}\ 0\ \bar{6}}$	$\simeq$ (S1.S2).(S2.S9).(S1.S9)
	$S_{\bar{3}\ \bar{9}\ \bar{5}}$	$\cdot S_{1\ 8\ 12}$	$\cdot S_{2\ 1\ \bar{7}}$	$\simeq$ (S2).(S9).(S2.S9)
	$S_{\bar{4}\ 8\ 6}$	$\cdot S_{\bar{1}\ \bar{8}\ \bar{12}}$	$\cdot S_{5\ 0\ 6}$	$\simeq$ (S1).(S9).(S1.S9)

Quartets

Again, as a consequence of the application of the correlation procedure, strengthened quartets play an important role in the initial stages of the sign generation procedure. Some strengthened quartets are given as an example in table 5.5. It is noted that the missing cross-term of the first quartet from table 5.5 is a reflection outside the measured reflection sphere, while the missing cross-terms of the next two quartets are reflections with $|E|$-values of 2.36 and 3.30 respectively.

Table 5.5 Strengthened quartets occuring in quadruple QI of table 5.4.

$S_{4\ \bar{8}\ \bar{6}} \cdot S_{3\ 9\ 5} \cdot S_{\bar{3}\ \bar{11}\ \bar{2}} \cdot S_{\bar{4}\ 10\ 3} \simeq +;\ \underline{h}+\underline{k} \equiv (7\ 1\ \bar{1});\ \underline{k}+\underline{l} \equiv (0\ 2\ \bar{3})$

$S_{4\ \bar{8}\ \bar{6}} \cdot S_{\bar{7}\ \bar{1}\ 1} \cdot S_{0\ \bar{2}\ 3} \cdot S_{3\ 11\ 2} \simeq +;\ \underline{h}+\underline{k} \equiv (\bar{3}\ \bar{9}\ \bar{5});\ \underline{h}+\underline{l} \equiv (4\ \bar{10}\ \bar{3})$

$S_{3\ 9\ 5} \cdot S_{\bar{7}\ \bar{1}\ 1} \cdot S_{0\ 2\ \bar{3}} \cdot S_{4\ \bar{10}\ \bar{3}} \simeq +;\ \underline{h}+\underline{k} \equiv (\bar{4}\ 8\ 6);\ \underline{h}+\underline{l} \equiv (3\ 11\ 2)$

Computer time

To perform the calculations dicussed in this chapter SCOR required about 7 minutes computer time on an IBM 370/158 computer. In each sign correlation cycle, 3000-6000 unique triples were considered. The sign determination was done very carefully in this example; in actual practice the sign determination would have been done much faster, with a corresponding decrease in computer time. For comparison: MULTAN (version of May 1971) required 4 minutes of computer time on the same machine to perform the sign determination for 32 solutions using the strongest 2000 unique triples.

CHAPTER 6

CALCULATION OF COSINE INVARIANTS

Structure invariants. The expressions $T_{\underline{h}_1\underline{h}_2} = S_{\underline{h}_1} \cdot S_{\underline{h}_2} \cdot S_{-\underline{h}_1-\underline{h}_2}$ and $\Phi_{\underline{h}_1\underline{h}_2} = \phi_{\underline{h}_1} + \phi_{\underline{h}_2} + \phi_{-\underline{h}_1-\underline{h}_2}$ play an important role in the theory and practice of direct methods. According to the sigma-2 relationship, $T_{\underline{h}_1\underline{h}_2} = +$ and $\Phi_{\underline{h}_1\underline{h}_2} = 0^\circ$ for triples with large triple products (see chapter 3). These expressions are said to be structure invariants, because they are independent of the choice of origin and are determined by the structure alone. That is, the value of $\Phi_{\underline{h}_1\underline{h}_2}$ is not affected by a change of origin.

There are many different forms of structure invariants. In theoretical discussions, a structure invariant is often restricted to a linear combination of phases. We then have: $\sum_{\underline{h}} c_{\underline{h}}\phi_{\underline{h}}$ is a structure invariant when $\sum_{\underline{h}} c_{\underline{h}} = 0$. Consequently, a triple:

$$\phi_{\underline{h}_1} + \phi_{\underline{h}_2} + \phi_{-\underline{h}_1-\underline{h}_2}$$

and a quartet:

$$\phi_{\underline{h}_1} + \phi_{\underline{h}_2} + \phi_{\underline{h}_3} + \phi_{-\underline{h}_1-\underline{h}_2-\underline{h}_3}$$

are examples of structure invariants. These expressions are nonetheless sensitive to a change of the enantiomorph, and - if anomalous dispersion is neglected - we cannot determine the enantiomorph from the experimental data.

The cosine of a structure invariant is said to be a cosine invariant and, in general, the values of cosine invariants are determined by the observed magnitudes of the normalized structure factors. Algebraic and probabilistic methods lead to formulae for the calculation of cosine invariants from the experimental data. The most intensively studied and hitherto most widely used cosine invariant is:

$$\cos \Phi_{\underline{h}_1\underline{h}_2} \equiv \cos(\phi_{\underline{h}_1} + \phi_{\underline{h}_2} + \phi_{-\underline{h}_1-\underline{h}_2}) \tag{6.1}$$

The expression:

$$|E_{\underline{h}_1}||E_{\underline{h}_2}||E_{\underline{h}_1+\underline{h}_2}| \cos \Phi_{\underline{h}_1\underline{h}_2} \tag{6.1a}$$

is called a triple invariant.

The B3,0 formula. The first formula reported for the calculation of the cosine invariant (6.1) is the so-called B3,0 formula (Hauptman and Karle, 1957, 1958; Karle and Hauptman, 1957,1958):

$$|E_{\underline{h}_1}||E_{\underline{h}_2}||E_{\underline{h}_1+\underline{h}_2}| \cos \phi_{\underline{h}_1\underline{h}_2} = AB_{\underline{h}_1\underline{h}_2} + C \qquad (6.2)$$

with A a positive constant, C a small positive correction term and

$$B_{\underline{h}_1\underline{h}_2} = \langle(|\xi_{\underline{k}}|^p - \overline{|\xi|^p})(|\xi_{\underline{h}_1+\underline{k}}|^q - \overline{|\xi|^q})(|\xi_{\underline{h}_1+\underline{h}_2+\underline{k}}|^r - \overline{|\xi|^r})\rangle_{\underline{k}} \qquad (6.3)$$

in which $\overline{|\xi|^p} = \langle|\xi_{\underline{k}}|^p\rangle_{\underline{k}}$. p , q and r are arbitrary real non-negative numbers. The averaging over $\underline{k}$ is done over as many reflections as possible, including the weak reflections. The B3,0 formula is frequently used with p, q and r values of 2. In this case eq. (6.3) becomes:

$$B_{\underline{h}_1\underline{h}_2} = \langle(|\xi_{\underline{k}}|^2 - 1)(|\xi_{\underline{h}_1+\underline{k}}|^2 - 1)(|\xi_{\underline{h}_1+\underline{h}_2+\underline{k}}|^2 - 1)\rangle_{\underline{k}} \qquad (6.4)$$

For triclinic space groups ($\epsilon_{\underline{h}} = 1$) this is:

$$B_{\underline{h}_1\underline{h}_2} = \langle(|E_{\underline{k}}|^2 - 1)(|E_{\underline{h}_1+\underline{k}}|^2 - 1)(|E_{\underline{h}_1+\underline{h}_2+\underline{k}}|^2 - 1)\rangle_{\underline{k}} \qquad (6.5)$$

The first calculations with the B3,0 formula were done in the early sixties (Karle, 1961; Karle and Karle, 1964). In subsequent years, several crystal structure determinations were carried out in which use was made of calculations with the B3,0 formula (Kanters et al., 1966; van Koningsveld, 1968; Kanters and Kroon, 1972a, 1972b; this thesis). Although the B3,0 formula does not permit the calculation of accurate cosine values, the formula proves to be useful when the interpretation of its results is done very carefully. In these applications, triple invariants were calculated for a relatively small number of triples with large triple products ($A_{\underline{h}_1\underline{h}_2}$; see eq. (3.8)). The results were used to find those triples which contradict the sigma-2 relationship i.e. triples with:

$$\phi_{\underline{h}_1} + \phi_{\underline{h}_2} + \phi_{-\underline{h}_1-\underline{h}_2} \approx 180^\circ \qquad (6.6)$$

(these triples will be called incorrect triples). The phase determination was carried out with the remaining triples (more details will be given in subsequent chapters).

In the applications just mentioned, the B3,0 formula was used to check the validity of the sigma-2 relationship for a number of strong triples. On the other hand, calculated cosine invariants can also be used explicitly. As noted before, the accuracy of the B3,0 formula is rather limited. Much effort has been given by Hauptman and coworkers to obtain formulae from which improved cosine values can be calculated.

The rather limited accuracy of the B3,0 formula was ascribed to the presence of overlap of peaks in the Patterson function (Hauptman, 1964). In order to decrease the influence of Patterson overlap on the results of the B3,0 formula some modifications were proposed: p, q and r were given the value 0.5, and the constant A was replaced by a sliding scale factor which is chosen in such a way that, for fixed triple product, the distribution of calculated cosine invariants agrees as closely as possible with the known theoretical distribution of cos $\Phi_{\underline{h}_1\underline{h}_2}$ (Hauptman et al., 1969). The theoretical probability distribution of cos $\Phi_{\underline{h}_1\underline{h}_2}$ is given for some values of $A_{\underline{h}_1\underline{h}_2}$ in fig. 6.1. In addition to the mentioned modifications, Hauptman proved in 1972 that it is not necessary to average over <u>all</u> reflections in eq. (6.3) but only over those reflections $\underline{k}$ for which $|E_{\underline{k}}|$ is larger than some threshold value, usually in the range 1.0 - 2.0 (Modified triple product formula or TPROD formula). In 1972, Hauptman also derived a complete new formula for the calculation of cos $\Phi_{\underline{h}_1\underline{h}_2}$ (MDKS formula). The presence of Patterson overlap has been taken into account in the derivation of this formula. It has been shown that, when extensive overlap is present in the Patterson synthesis, the computed cosine values from the MDKS formula are more accurate than from the TPROD formula (Duax et al., 1972).

In the first structure determinations using cosine invariants calculated with the TPROD and/or MDKS formulae (e.g. Hauptman et al., 1969; Weeks et al., 1971) a procedure was used consisting of the following steps:

a) calculation of some one to two thousand cosine invariants,

b) determination of the phases of some one hundred reflections from the calculated cosine invariants by a least-squares procedure, and

c) refinement and extension of the phases with the tangent formula.

This procedure is usually called the least-squares analysis of cosine

invariants.

It proved that the accuracy with which the cosine invariants can be calculated with the TPROD and MDKS formulae is still poor. Several techniques were used by Hauptman and coworkers to make most effective and safest use of the calculated cosine invariants.

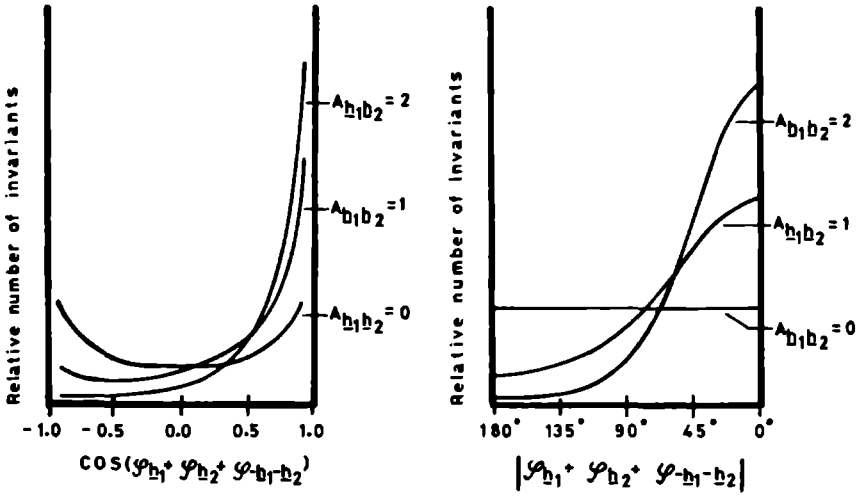


Fig. 6.1 Theoretical probability distributions of $\cos(\phi_{\underline{h}_1}+\phi_{\underline{h}_2}+\phi_{-\underline{h}_1-\underline{h}_2})$ *and* $(\phi_{\underline{h}_1}+\phi_{\underline{h}_2}+\phi_{-\underline{h}_1-\underline{h}_2})$ *for some values of* $A_{\underline{h}_1\underline{h}_2}$ *(see Hauptman, 1972a; eqs. (8.3) and (9.3) of chapter III).*

References

Duax, W.L., Weeks, C.M. and Hauptman, H. (1972). Acta Cryst. B28, 1857.

Hauptman, H. (1964). Acta Cryst. 17, 1421.

Hauptman, H. (1972). Z. Kristallogr. 155, 1.

Hauptman, H. (1972a). Crystal Structure Determination. The Role of the Cosine Seminvariants. Plenum Press. New York.

Hauptman, H., Fisher, J., Hancock, H. and Norton, D.A. (1969). Acta Cryst. B25, 811.

Hauptman, H. and Karle, J. (1957). Acta Cryst. 10, 267.

Hauptman, H. and Karle, J. (1958). Acta Cryst. 11, 149.

Karle, I.L. (1961). Acta Cryst. 14, 497.

Karle, I.L. and Karle, J. (1964). Acta Cryst. 17, 835.

Karle, J. and Hauptman, H. (1957). Acta Cryst. 10, 515.

Karle, J. and Hauptman, H. (1958). Acta Cryst. 11, 264.

Kanters, J.A. and Kroon, J. (1972a). Acta Cryst. B28. 1345.

Kanters, J.A. and Kroon, J. (1972b). Acta Cryst. B28, 1946.

Kanters, J.A., Kroon, J., Beurskens, P.T. and Vliegenthart, J.A. (1966). Acta Cryst. 21, 990.

van Koningsveld, H. (1968). Recueil 87, 243.

Weeks, C.M., Cooper, A., Norton, D.A., Hauptman, H. and Fisher, J. (1971). Acta Cryst. B27, 1562.

CHAPTER 7

THE USE OF CALCULATED TRIPLE INVARIANTS

7.1 The general idea

The sigma-2 procedures, based on the application of the sum of angle formula and analogous formulae (see chapter 3) often lead to the correct solution of the phase problem. Sometimes they fail, often because of incorrect triples (see eq. (6.6)) that have been used during the initial steps of the phase determination. By variations in execution parameters {such as selection criteria, origin and ambiguity choices}, these incorrect triples may accidently be avoided, and the correct solution of the problem obtained.

On the other hand, structures can be solved by using cosine invariants, calculated from experimental data. Hauptman and coworkers used the least-squares analysis of calculated cosine invariants (chapter 6). In this case, the cosine values are used, as they are calculated from experimental data, rather than the cosine being equated to +1 (the sigma-2 procedures). The procedure based on the calculated cosine invariants is more powerful than each of the sigma-procedures, but requires more computer time. In addition, the quality of the calculated cosine invariants is not good enough to ensure success in all cases.

Another procedure was also mentioned in chapter 6. In that procedure, sigma-2 procedures are strengthened by preliminary calculations of a limited number of triple invariants, at the cost of a small increase in computer time. The calculated invariants are used to check the validity of the sigma-2 relationship. The calculation of even a small number of invariants gives useful results. Accurate cosine values are not required: the results of the invariant calculations are merely used to reject the triples which have a large chance of being incorrect, and to select the triples which should be used in the early stages of the phasing process. Application of a sigma-2 procedure to the triples, from which the majority of incorrect triples have been withdrawn, increases the chances of finding the correct solution to the phase problem. When a relatively large number of the strongest reflections (say, one hundred) have been phased in this way, the phase determination may be continued by routine sigma-2 procedures.

We perform the triple invariant calculations with the B3,0 formula, using eqs. (6.2) and (6.4). The numerical results of the calculations with the B3,0 formula (on seven structures) show a wide range of $AB_{\underline{hk}} + C$ values; the results for triples with large $A_{\underline{hk}}$ values can be interpreted in the following way:

a. $AB_{\underline{hk}} + C \gg 0$. Very large values are reliable indications for $\cos \Phi_{\underline{hk}} = +1$, thus affirming the validity of the sigma-2 relationship. Medium large $AB_{\underline{hk}} + C$ values are still good affirmations of the validity of the sigma-2 relationship.

b. $AB_{\underline{hk}} + C \approx 0$. Weak results are unreliable. For non-centric structures these weak results are certainly no good indications for $\cos \Phi_{\underline{hk}} \approx 0$. The theoretical distribution curve of $\cos \Phi_{\underline{hk}}$ (fig. 6.1) does have a minimum where the experimental $AB_{\underline{hk}} + C$ distribution shows a maximum. For centric structures $\cos \Phi_{\underline{hk}}$ can only have the values ±1. Many of the weak results correspond to $\cos \Phi_{\underline{hk}} = +1$.

c. $AB_{\underline{hk}} + C \ll 0$. Very large negative results are strong indications that the sigma-2 relationship is not satisfied. The possible implication: $\cos \Phi_{\underline{hk}} = -1$ is of limited importance, i.e. the reliability of the indication $\cos \Phi_{\underline{hk}} = -1$ is not large enough to be of practical use.

It is clear that the triples with the highest $AB_{\underline{hk}} + C$ values will be used in the initial steps of the phase determination with a sigma-2 procedure. As the phase determination progresses, more and more triples with lower $AB_{\underline{hk}} + C$ values may enter into the procedure. Despite the fact that triples with very high $AB_{\underline{hk}} + C$ values are very reliable, incorrect triples may occur, occasionally. Therefore, we prefer to use the phase correlation procedure to determine the phases of the individual reflections.

7.2 Applications of the procedure

It is in general much more difficult to solve a non-centric structure than a centric one. For centric structures $\cos \Phi_{\underline{hk}}$ can have only the values +1 and −1; for non-centric structures $\cos \Phi_{\underline{hk}}$ may have any value between +1 and −1, for general reflections. In case a non-centric structure has one or more centric projections, it will be of great help to have the

phases of a number of projection reflections before extending the phase determination to general (three-dimensional) reflections. A centric projection can be solved relatively easily (a projection is considered to be solved in case the signs of the projection reflections are expressed in terms of a few unknown letter symbols). In solving a projection, not only triples among projection reflections are used, but also triples in which a projection reflection and two symmetry-related three-dimensional reflections occur; we will call these triples: sigma-1 type sigma-2 interactions.

These considerations lead to the classification of space groups in four categories:

1. Centric space groups. In this case triple invariants are calculated with the B3,0 formula for all triples that can be generated from all strong reflections (say, with $|E|>1.7$). Thereafter the sign correlation procedure is applied to the triples with the largest $AB_{\underline{hk}}+C$ values. Further sign generation is done with routine sigma-2 procedures; triples having negative $AB_{\underline{hk}}+C$ values are excluded. Applications of this procedure to the crystal structure determinations of 'LAAR' and 'HEPTA.C' will be included in this thesis (chapters 8 and 17, respectively).

2. Non-centric space groups with three centric projections, e.g. $P2_12_12_1$. Apply the B3,0 formula to all triples among projection reflections with, say, $|E|>1.0$ (also included are triples between reflections from different projections (triples with, by symmetry, $\Phi_{\underline{hk}} = \pm 90^{\circ}$ are omitted)), and also to triples of the sigma-1 type; then the projections are expressed in terms of a few letter symbols, constituting a very good starting set for further phase generation (of three-dimensional reflections) by routine applications of the sum of angle formula and related formulae. The use of the phase correlation procedure followed by tangent formula refinement is advised. As an example of this category is the crystal structure determination of 'DIOX' given in chapter 9.

3. Non-centric space groups with one centric projection, e.g. $P2_1$. The centric projection may be solved in the same way as described in 2.; in this case it makes sense also to use information from other sources, such as from the sigma-3 and pair relationships (see chapter 10). Phase extension to general reflections is done as described in 1., using phase correlation on results from the B3,0 formula. Again, results from other sources may be used. The phase generation is then continued by routine

applications of the sum of angle formula, followed by tangent formula refinement. The crystal structure determination of 'HEPTA.NC', in the space group $P2_1$, is described in chapter 11.

4. Non-centric space groups without a centric projection (space groups which have only centric axial reflections also belong to this category), e.g. P1, Pm. Our procedure has never been tried in such space groups.

7.3 Calculated triple invariants in automatic phase determination procedures

In this section we will discuss a weighting scheme from which the reliability of triples, for which triple invariants have been calculated, can be estimated, relative to the reliability of the sigma-2 relationship. The proposed scheme will be incorporated in our NIJDIR system of automatic direct methods computer programs (appendix A_1). This scheme has been developed after inspection of the results for several crystal structures (table 7.1). To illustrate the arguments and the discussion in this

Table 7.1 Structures which have been used for triple invariant calculations with the B3,0 formula.

Structure	Formula	Sp. Gr.	Z	References
'LAAR'	$C_{44}H_{36}$	C2/c	4	Chapter 8 of this thesis van den Hark and Beurskens,1973 van den Hark, 1974b
'DIOX'	$C_{18}O_4H_{20}$	$P2_12_12_1$	4	Chapter 9 of this thesis van den Hark and Beurskens,1975
'HEPTA.NC'	$C_{30}H_{18}$	$P2_1$	4	Chapter 11 of this thesis van den Hark, 1976a
'HEPTA.C'	$C_{30}H_{18}$	$P2_1/c$	4	Chapter 17 of this thesis
Raffinose pentahydrate	$C_{18}H_{32}O_{16}.5H_2O$	$P2_12_12_1$	4	Berman, 1970 van den Hark, 1974a van den Hark, 1976b
Δ-8,14-Anhydrodigitoxigenin	$C_{23}H_{32}O_3$	$P2_12_12_1$	4	Gilardi and Karle, 1970 van den Hark, 1974a van den Hark, 1976b
E-mesithyl(phenylsulphinyl)sulphine	$C_{16}H_{16}O_2S_2$	$P2_1/c$	4	Hummelink, 1974 van den Hark, 1973

section we will use numerical results for one of these structures: 'DIOX'. For this structure, triple invariants were calculated for all 1166 triples between projection reflections with $|E|>1.0$ (triples with $\Phi_{\underline{hk}} = \pm 90^{o}$ have been omitted) and for all 557 triples between general reflections with $|E|>1.55$.

In order to define the weighting scheme, we will introduce $KAB_{\underline{hk}}$ values which are related to $AB_{\underline{hk}}$ + C values. We will first define $KAB_{\underline{hk}}$ values and summarize some characteristics; additional information will be given later in this section.

Definition of $KAB_{\underline{hk}}$ values. The $AB_{\underline{hk}}$ + C values, as they are calculated with the B3,0 formula, show a wide range of values, usually quite different for different structures. Therefore, comparison of the results of the B3,0 formula in terms of $AB_{\underline{hk}}$ + C values is difficult. Define:

$$KAB_{\underline{hk}} = K (AB_{\underline{hk}} + C + X) \qquad (7.1)$$

with K a scaling factor and X a constant which are adjusted to obtain the following characteristics:

1. the order of the triples is the same for sorting on $AB_{\underline{hk}}$ + C values as for sorting on $KAB_{\underline{hk}}$ values.
2. negative $KAB_{\underline{hk}}$ values indicate unreliable sigma-2 interactions.
3. positive $KAB_{\underline{hk}}$ values have magnitudes which are comparable with $A_{\underline{hk}}$ values; this permits the estimation of the reliability; if $KAB_{\underline{hk}} > A_{\underline{hk}}$ the reliability of the triple is larger than that of the corresponding sigma-2 interaction (e.g. eq. (3.3)).
4. $KAB_{\underline{hk}}$ values for triples from different structures are comparable (because of 3.).

It is important for automatic phase determination procedures to have a proper scheme from which the reliability of a triple can be estimated. In chapter 3 we have seen that in the sigma-2 procedures the reliability of a triple is related to its $A_{\underline{hk}}$ value. In general, the reliability of a triple will decrease for decreasing $A_{\underline{hk}}$, i.e. for decreasing $A_{\underline{hk}}$ an increasing number of triples do not satisfy the sigma-2 relationship. This illustrated in table 7.2: for triples from projection reflections the relative number of incorrect triples ($\Phi_{\underline{hk}} = 180^{o}$) increases with decreasing $A_{\underline{hk}}$; for triples from general reflections the relative number of triples with $|\Phi_{\underline{hk}}| > 90^{o}$ (which we also call incorrect triples) increases.

It was noted in section 7.1 that the reliability of calculated triple invariants increases with increasing $AB_{\underline{hk}} + C$. This is illustrated in table 7.3. The results are given for ranges of $KAB_{\underline{hk}}$. Note that only few triples with large $KAB_{\underline{hk}}$ values are incorrect (again, incorrect with respect to the sigma-2 relationship); on the other hand, many triples with negative $KAB_{\underline{hk}}$ values are incorrect. On comparing table 7.3 with table 7.2, it is clear that the reliability of triples with large $KAB_{\underline{hk}}$ values is much better than with large $A_{\underline{hk}}$ values (say, $KAB_{\underline{hk}}$ and $A_{\underline{hk}}$ values larger than 2.0). Moreover, there are more triples with large $KAB_{\underline{hk}}$ values than with large $A_{\underline{hk}}$ values.

It can be expected that the reliability of a triple, which is dependent on its $KAB_{\underline{hk}}$ value, should also be a function of its $A_{\underline{hk}}$ value. It proves that the reliability of a triple with a given $KAB_{\underline{hk}}$ value, in general, is larger when its $A_{\underline{hk}}$ value is larger. This can be seen from figure 7.1 where the reliability is plotted as a function of $KAB_{\underline{hk}}$ for two different ranges of $A_{\underline{hk}}$.

Table 7.2 The reliability of triples for different ranges of $A_{\underline{hk}}$ for 'DIOX' ($A_{\underline{hk}} = 0.212|E_h||E_k||E_{\underline{h+k}}|$). The phases of the reflections were calculated from the refined atomic positions.

	Triples from projection reflections		Triples from general reflections		
Range of $A_{\underline{hk}}$	Total	Incorrect	Total	Incorrect	$<\|\phi_{\underline{hk}}\|>$
> 2.00	14	7 %	56	7 %	37.5°
1.50 - 2.00	46	17	118	6	41.3
1.25 - 1.50	61	13	161	20	53.4
1.00 - 1.25	145	21	171	18	55.4
0.80 - 1.00	190	26	51	24	56.3
0.60 - 0.80	287	30			
0.40 - 0.60	331	39			
0.20 - 0.40	92	41			

Table 7.3 The reliability of triples for different ranges of $KAB_{\underline{hk}}$ for 'DIOX'.

	Triples from projection reflections		Triples from general reflections		
Range of $KAB_{\underline{hk}}$	Total	Incorrect	Total	Incorrect	$<\lvert\phi_{\underline{hk}}\rvert>$
> 2.00	103	2 %	114	3 %	30.7°
1.50 - 2.00	114	10	61	7	38.4
1.00 - 1.50	158	16	93	3	37.7
0.50 - 1.00	214	21	104	12	49.6
0.00 - 0.50	212	33	88	24	60.9
< 0.00	365	54	97	44	83.3

Table 7.4 The reliability of triples for different ranges of $W_{\underline{hk}}$ for 'DIOX'.

	Triples from projection reflections	
Range of $W_{\underline{hk}}$	Total	Incorrect
> 2.00	99	4 %
1.50 - 2.00	50	4
1.00 - 1.50	110	5
0.50 - 1.00	220	18
0.00 - 0.50	322	32
0.0	365	54

As noted before, we calculate triple invariants for a limited number of triples, to save computer time. Thus it is necessary that the scheme, from which the reliability of a triple is estimated, permits the simultaneous use of triples for which triple invariants have been calculated and triples for which this has not been done.

Note: We prefer the calculation of triple invariants to the calculation of cosine invariants. When the triple invariant is divided by the corresponding value of $|E_{\underline{h}}||E_{\underline{k}}||E_{\underline{h+k}}|$ the cosine invariant is obtained. For triples with small $A_{\underline{hk}}$ values, the calculated triple invariant proved

$0.78 < A_{\underline{hk}} < 1.25$ $0.26 < A_{\underline{hk}} < 0.56$

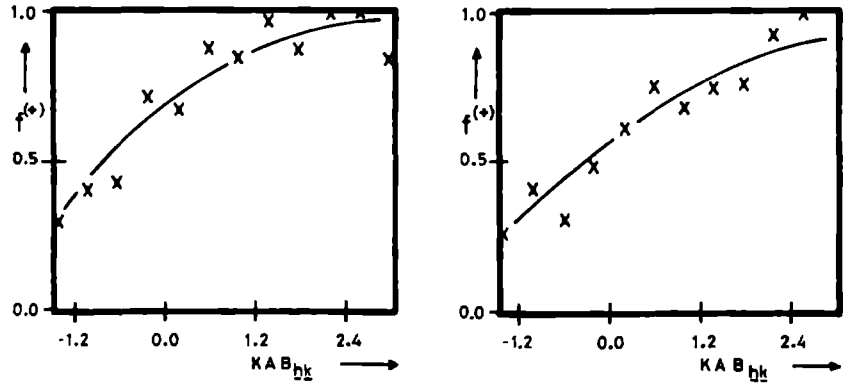


Fig. 7.1 The fraction of correct triples ($f^{(+)}$) as a function of $KAB_{\underline{hk}}$, for two different ranges of $A_{\underline{hk}}$ (only for triples from projection reflections).

to be too optimistic an estimate for the reliability, in comparison with triples with the same triple invariant but with larger $A_{\underline{hk}}$ values (fig. 7.1). The cosine invariants will exaggerate this effect.

$KAB_{\underline{hk}}$ values are calculated from the $AB_{\underline{hk}}$ + C values with eq. (7.1); now we will describe the definition and calculation of K and X in some more detail. The arguments are illustrated using the triples from projection reflections of 'DIOX'.

The constant X is added to the calculated $AB_{\underline{hk}}$ + C values to achieve that triples with a positive $KAB_{\underline{hk}}$ value have a probability of being correct of at least 50% (i.e. $|\Phi_{\underline{hk}}| < 90^{o}$). The constant X can be chosen in such a way that, for fixed $A_{\underline{hk}}$, the fraction of triples with negative $KAB_{\underline{hk}}$ values agrees with the fraction of negative triple invariants that is expected theoretically (denoted by $f^{(-)}$). For triples from centric structures or centric projections of non-centric structures, $f^{(-)}$ is given by (using eq. (3.3)):

$$f^{(-)} = 0.5 - 0.5 \tanh(0.5\, A_{\underline{hk}}) \qquad (7.2)$$

For triples from general reflections of non-centric structures, values of $f^{(-)}$ are tabulated for different $A_{\underline{hk}}$ values by Hauptman (1972).

Let us sort the triples for which triple invariants have been calculated in order of decreasing $A_{\underline{hk}}$ and let us divide them in ten groups containing equal numbers of triples. In each of the groups, $A_{\underline{hk}}$ is essentially constant. Let us define:

$N^{(t)}$: the number of triples in each of the groups.

$N^{(-)}$: the expected number of incorrect triples in each of the groups, $N^{(-)} = f^{(-)} N^{(t)}$.

$N^{(+)}$: the expected number of correct triples in each of the groups, $N^{(+)} = N^{(t)} - N^{(-)}$.

$(AB_{\underline{hk}} + C)^{\pm}$: the cut-off $AB_{\underline{hk}} + C$ value in each of the groups (i.e. the largest $AB_{\underline{hk}} + C$ value of the $N^{(-)}$ triples with the smallest $AB_{\underline{hk}} + C$ values).

$N_{inc}^{(-)}$: the actual number of incorrect triples among the $N^{(-)}$ triples with the smallest $AB_{\underline{hk}} + C$ values.

$N_{inc}^{(+)}$: the actual number of incorrect triples among the $N^{(+)}$ triples with the largest $AB_{\underline{hk}} + C$ values.

Then it can be expected that: $N_{inc}^{(-)} + N_{inc}^{(+)} \approx N^{(-)}$.

Some results for triples from projection reflections of 'DIOX' are given in table 7.5. The $AB_{\underline{hk}} + C$ values are in the range: -61.5 to 138.6. It can be seen from table 7.5 that $(AB_{\underline{hk}} + C)^{\pm}$ is approximately independent of $A_{\underline{hk}}$. Therefore, we use a constant value for X, for all triple invariants, which is calculated from the average of the individual cut-off values or from the cut-off value for all calculated triple invariants:

$$X = -(AB_{\underline{hk}} + C)^{\pm}_{\text{all triples}} \qquad (7.3)$$

It should be stressed once more that only the results for 'DIOX' are discussed here; the results for other structures, given in table 7.1, lead to the same conclusions.

The scaling factor K is chosen such that the positive $KAB_{\underline{hk}}$ values are comparable with the $A_{\underline{hk}}$ values. Define E_x as the theoretical expectation value of $(A_{\underline{hk}} \cos \Phi_{\underline{hk}})$. For triples from centric structures or from centric projections of non-centric structures, with fixed $A_{\underline{hk}}$ values,

Table 7.5 The $AB_{\underline{hk}}$+ C cut-off value for different ranges of $A_{\underline{hk}}$.

Range of $A_{\underline{hk}}$	$\langle A_{\underline{hk}} \rangle$	$f^{(-)}$	$N^{(+)}$	$N^{(-)}$	$(AB_{\underline{hk}} + C)^{(\pm)}$	$N^{(+)}_{inc}$	$N^{(-)}_{inc}$
2.29 - 1.25	1.57	0.17	97	20	-8.2	8	8
1.25 - 1.04	1.13	0.24	89	28	-4.7	7	12
1.04 - 0.89	0.95	0.28	85	32	-6.3	14	18
0.89 - 0.78	0.84	0.30	82	35	-9.2	13	19
0.78 - 0.71	0.74	0.32	80	37	-4.4	12	19
0.71 - 0.62	0.67	0.34	78	39	-6.3	16	21
0.62 - 0.56	0.59	0.36	76	41	-4.7	21	26
0.56 - 0.49	0.52	0.37	74	43	-7.8	24	21
0.49 - 0.40	0.45	0.39	72	45	-3.6	15	32
0.40 - 0.26	0.35	0.41	66	47	-3.1	19	25
2.29 - 0.26	0.78	0.31	799	367	-5.7	149	201
					X = 5.7		

E_x is given by (from eq. (7.2)):

$$E_x = A_{\underline{hk}} (1.0 - 2f^{(-)}) \tag{7.4}$$

For triples from general reflections of non-centric structures, an expression for the calculation of E_x, as a function of $A_{\underline{hk}}$, is given by Hauptman (1972). The scaling of the triple invariants can be done in two ways:

1. with a 'sliding' scaling factor, K_S, which is a function of $A_{\underline{hk}}$. K_S can be calculated from:

$$K_S = E_x / \langle AB_{\underline{hk}} + C + X \rangle \tag{7.5}$$

for triples with fixed $A_{\underline{hk}}$ values. It happens that K_S increases approximately linearly with increasing $A_{\underline{hk}}$ (fig. 7.2). From this, it can also be shown that $\langle AB_{\underline{hk}} + C + X \rangle$ increases approximately linearly with increasing $A_{\underline{hk}}$ (this approximation is not valid for extremely large $A_{\underline{hk}}$ values). This linear relationship is used in the definition of the weighting scheme (see below).

2. with a constant scaling factor, K.

When a constant scaling factor is used for the calculation of $KAB_{\underline{hk}}$

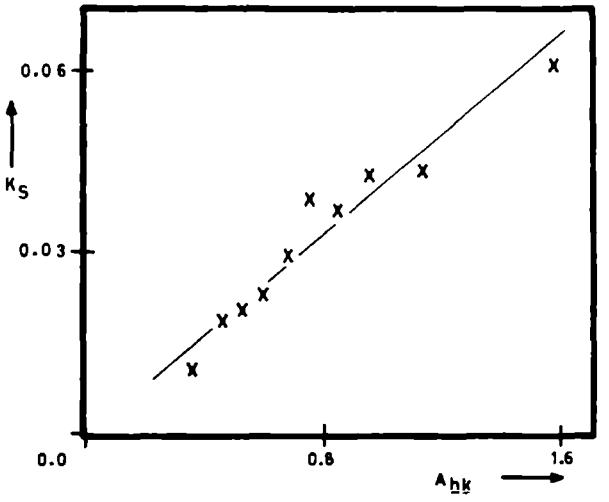


Fig. 7.2 The 'sliding' scaling factor, K_S, as a function of A_{hk} for triples from projection reflections of 'DIOX'.

values (eq. (7.1)), the KAB_{hk} values are determined completely by the results of the triple invariant calculations. We prefer the use of a constant scaling factor, calculated from eq. (7.5) using some 50 to 100 triples with the largest A_{hk} values. In that case, as estimates of the reliability of the triples, the KAB_{hk} values for triples with small A_{hk} values will be too large, in comparison with the KAB_{hk} values for triples with larger A_{hk} values (see fig. 7.1). In the expression for the weight, from which the reliability of a triple is estimated (see below), this fact will be taken into account.
Note. The KAB_{hk} values given in this section for 'DIOX' were calculated with a scaling factor K = 0.061.

The weights W_{hk}. We assign weights, W_{hk}, to all triples that are to be used for phase determination, that is: both to triples for which triple invariants have been calculated and to triples for which such calculations have not been done. We will define a weighting scheme such that all triples can be used simultaneously for phase determination. The reliability of a triple can be estimated with existing formulae, such as eqs. (3.3) and (3.7); we define the weights such that in these equations A_{hk} can be replaced by W_{hk}.

Triples for which no triple invariants have been calculated, are therefore weighted by:

$$W_{\underline{hk}} = A_{\underline{hk}} \tag{7.6}$$

Triples for which triple invariant calculations resulted in negative $KAB_{\underline{hk}}$ values will not be used for phase determination and therefore:

$$W_{\underline{hk}} = 0 \tag{7.7}$$

Triples for which triple invariant calculations resulted in positive $KAB_{\underline{hk}}$ values are weighted by:

$$W_{\underline{hk}} = 2\,(A_{\underline{hk}} \cdot KAB_{\underline{hk}})^{\frac{1}{2}} \tag{7.8}$$

The reasons for this definition are as follows:

- As a consequence of the scaling procedure, the triples with the largest $A_{\underline{hk}}$ values will have $KAB_{\underline{hk}}$ values which, on the average, are comparable with their $A_{\underline{hk}}$ values, for predicting the reliability of the triples.
- For triples with $A_{\underline{hk}} = KAB_{\underline{hk}}$, eq. (7.8) leads to double the weight given by eq. (7.6).
- For the strongest triples, the top results for $KAB_{\underline{hk}}$ will have maximum weight; extremely large $KAB_{\underline{hk}}$ values will be reduced by the square root in eq. (7.8).
- As shown in figs. 7.1 and 7.2, triples with smaller $A_{\underline{hk}}$ values will, on the average, have less reliable $KAB_{\underline{hk}}$ values. As we did not use the 'sliding' scale, K_S, the $KAB_{\underline{hk}}$ values for smaller $A_{\underline{hk}}$ values give an overestimation for the weight, and therefore the product of $A_{\underline{hk}}$ and $KAB_{\underline{hk}}$ appears in eq. (7.8).

In table 7.4 are the results of the triple invariant calculations for 'DIOX' given in terms of weights. It is clear that the triples which play a key role in the phase determination, that is triples with large weights (say, larger than 1.0) are reliable; as the weight becomes smaller, the reliability drops very rapidly.

A comparison of tables 7.2, 7.3 and 7.4 shows that the calculation of triple invariants and the weighting of the triples as is done with

eq. (7.8) is useful. For example:

in table 7.2, 47 of the 266 triples with $A_{\underline{hk}}$ > 1.0 are incorrect

in table 7.3, 39 of the 375 triples with $KAB_{\underline{hk}}$ > 1.0 are incorrect

in table 7.4, 11 of the 259 triples with $W_{\underline{hk}}$ > 1.0 are incorrect.

With the weighting scheme discussed in this section we now are able to introduce calculated triple invariants in automatic phase determination procedures.

7.4 References

Berman, H. (1970). Acta Cryst. B26, 290.

Gilardi, R.D. and Karle, I.L. (1970). Acta Cryst. B26, 207.

van den Hark, Th.E.M. (1973). Werkrapport, Nr. 2, Crystallography Laboratory, University of Nijmegen.

van den Hark, Th.E.M. (1974a). Werkrapport, Nr. 7, Crystallography Laboratory, University of Nijmegen.

van den Hark, Th.E.M. (1974b). Direct Methods Meeting, Utrecht.

van den Hark, Th.E.M. (1976a). Werkrapport, Nr. 14, Crystallography Laboratory, University of Nijmegen.

van den Hark, Th.E.M. (1976b). Werkrapport, Nr. 15, Crystallography Laboratory, University of Nijmegen.

van den Hark, Th.E.M. and Beurskens, P.T. (1973). Abstract, Hauptman Seminar, Parma.

van den Hark, Th.E.M. and Beurskens, P.T. (1975). Direct Methods Meeting, Utrecht.

Hauptman, H. (1972). Crystal Structure Determination. The Role of the Cosine Seminvariants. Plenum Press. New York.

Hummelink, Th.W. (1974). J. Cryst. Mol. Struct. 4, 373.

CRYSTAL STRUCTURE OF 'LAAR'

Reprinted from J. Cryst. Mol. Struct. (1974) 4, 227 - 242 (the table with structure factors has been omitted for brevity).

Application of the *B*2,0 and *B*3,0 formulae to the structure determination of a photodimer of *o*-distyrylbenzene, $C_{44}H_{36}$

TH. E. M. VAN DEN HARK, P. T. BEURSKENS AND W. H. LAARHOVEN

Crystallography Laboratory, and Department of Organic Chemistry, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands

(Received 20 August 1973)

Abstract

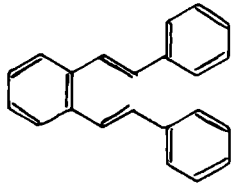
A simple application of the *B*3,0 formula is described. This formula is used mainly to avoid inconsistent Σ_2-interactions in standard symbolic-addition or multiple-solution techniques. Computer time is reduced by a reflection selection procedure based upon the use of the *B*2,0 formula. The solution of a centrosymmetric structure is described. This compound, a photodimerization product of *o*-distyrylbenzene, is 5,6,11,12-tetraphenyl-dibenzo[2-3,8-9]tricyclo[8,2,0,$0^{4,7}$]dodecadiene-2,8, $C_{44}H_{36}$. The compound crystallizes in the monoclinic space group *C*2/*c*, with unit cell parameters $a = 28.047$, $b = 9.504$, $c = 12.600$ Å and $\beta = 103.4°$. A rather poor set of data was collected by an automatic diffractometer. Structural parameters were refined by full-matrix least-squares methods to an *R*-value of 0.06 for 970 non-zero reflections. The molecule is situated on a twofold rotation axis. It contains *cis*-, *trans*-, *cis*-substituted puckered cyclobutane rings. The dihedral angle between the benzo-groups is 61.5°.

Introduction

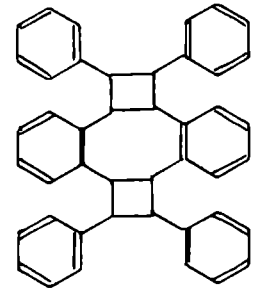
Attention is given to the solution of a phase problem that, at first sight, may be considered as easily solvable, namely a small, monoclinic, centro-

symmetric, light-atom structure. Standard techniques, however, failed for this example: the Patterson synthesis could not be solved because of heavy overlap and lack of resolution; attempts to solve the structure by standard multi-solution and symbolic-addition procedures failed because of inconsistent Σ_2-relations between reflections with large E-values and, probably, because of the small number of reflections available. Certainly, the structure could have been solved by repeated calculations, on modifying the computer input parameters, and also by modern probabilistic techniques (Duax et al, 1972). Nevertheless, we assume that it will be of interest to develop a simple procedure that may be used routinely for small structures when standard techniques fail at a first trial. The procedure, used to solve the present structure, is easily programmed and does not require much computer time.

On irradiation of *o*-distyrylbenzene (I) several compounds are formed. Among them, three dimeric molecules, formed by a twofold cyclization, are present. These compounds were reported independently by Müller et al (1966, 1970) and by Laarhoven et al (1970), and have the same m.p., u.v. and n.m.r. data. However, different molecular structures were assigned to these products. Müller et al (1966, 1970) proposed that these dimers were isomers of structure (II), whereas Laarhoven et al (1970) gave structures without cyclobutane rings, because the expected formation of stilbene on pyrolysis of compounds like (II) did not occur. Therefore, an X-ray analysis was carried out of one of the dimeric products (m.p. 293°).



(I)



(II)

Experimental

5,6,11,12-tetraphenyl-dibenzo[2-3,8-9] tricyclo [8,2,0,$0^{4,7}$]dodecadiene-2,8 ($C_{44}H_{36}$), FW = 564.87, forms small, colourless crystals, somewhat elongated along the *b*-axis.

The crystals are monoclinic with space group $C2/c$ (No. 15). From Pt-calibrated Weissenberg photographs taken with Ni-filtered Cu $K\alpha$ radiation (λ = 1.5418 Å), application of a least-squares procedure yielded a = 28.047(9), b = 9.504(3), c = 12.600(4) Å and β = 103.4(1)°;

$V_c = 3267(3)$ Å^3. The calculated density of 1.148 g cm^{-3} with $Z = 4$, agrees with the measured value of 1.13 g cm^{-3} (flotation method); $F(000) = 1200$.

Intensity data were collected with an automatic NONIUS diffractometer, using Zr-filtered Mo $K\alpha$ radiation up to a $\sin\theta/\lambda$-value of 0.48. Of the 1519 attainable symmetry-independent reflections, 970 were observed above background ($I \geqslant 3\sigma$). After every 20 reflections, a reference reflection was measured to detect and allow corrections to be made for slow fluctuations in the primary beam.

Corrections were made for Lorentz and polarization factors, the data were placed on an absolute scale by means of a Wilson plot and normalized structure factors were calculated. The experimental distribution of normalized structure factors is consistent with the centrosymmetric space group $C2/c$.

Structure determination

B3,0 and B2,0 formulae. The $B3,0$ formula (Hauptman & Karle, 1957, 1958; Karle & Hauptman, 1957, 1958, 1959) is written as

$$E'_{\mathbf{h}} E'_{\mathbf{h}'} E'_{-\mathbf{h}-\mathbf{h}'} = AB + C \qquad (1)$$

with

$$B = \langle (|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{k}+\mathbf{h}}|^2 - 1)(|E_{\mathbf{k}+\mathbf{h}+\mathbf{h}'}|^2 - 1) \rangle_{\mathbf{k}}$$

in which the average is taken over *all* reflections $\mathbf{k} = hkl$; A is a positive constant and C is a small positive correction term. E' is the normalized structure factor of the 'squared structure'. For *large* $|E|$-values, the approximation $E' \approx E$ may be used. According to our experience (Kanters et al, 1966), the numerical results are rather poor. About 50% of the results, however, may be used in a sign generating procedure: the sign of the left hand side of (1) is equal to the calculated sign of $AB + C$ provided that $|AB + C| \gg 0$:

$$S(E_{\mathbf{h}} E_{\mathbf{h}'} E_{-\mathbf{h}-\mathbf{h}'}) = S(AB + C) \qquad (2)$$

Equation (2) can be used only for large values of the triple product $|E_{\mathbf{h}} E_{\mathbf{h}'} E_{-\mathbf{h}-\mathbf{h}'}|$, otherwise large deviations in the $B3,0$ results are to be expected. It may be noted that in the usual application of the Σ_2-relation, the left-hand side of (2) is supposed to be positive; the $B3,0$ formula provides a possibility to avoid inconsistent (non-valid) interactions.

The sigma-1 type $B3,0$ formulae are obtained by using symmetry-related reflections $\mathbf{h}$ and $\mathbf{h}'$; for space group $C2/c$, with $E \approx E'$:

$$\begin{aligned} (-1)^l |E_{hkl}|^2 E_{0,2k,0} &= AB + C \\ (-1)^l |E_{hkl}|^2 E_{2h,0,2l} &= AB + C \end{aligned} \qquad (3)$$

The $B2,0$ formula (Karle & Hauptman, 1959) is written as

$$|E'_{\mathbf{h}}|^2 = 1 + D\langle(|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{k}+\mathbf{h}}|^2 - 1)\rangle_{\mathbf{k}} \qquad (4)$$

in which D is a positive constant. This formula is used to calculate the $|E'|$-values of the medium-strong reflections. Where a medium-strong reflection is used in the triple product equation (1), and if $|E'| \ll |E|$ for this reflection, many of the $B3,0$ results are likely to be unreliable. So the $B2,0$ formula is applied to select reflections in order to enhance the probability of finding useful results (large $(AB + C)$-values).

Calculations. The $B3,0$ formula was used to calculate $(AB + C)$-values for 460 interactions $(\mathbf{h},\mathbf{h}',\mathbf{h}+\mathbf{h}')$ among 114 reflections with $|E| \geqslant 1.7$ (table 1a). The $B2,0$ formula (4) was used to calculate $|E'|$-values of the 90 reflections with $1.4 < |E| < 1.7$. For 26 reflections, we found $|E'|$ to be greater than 2.5.

Including these reflections, the $B3,0$ calculations were extended to 332 additional interactions (table 1b). The sigma-1 type $B3,0$ formula (3) was applied to $0,2k,0$ and $2h,0,2l$ reflections with $|E| \geqslant 1.0$.

Table 1. *Distribution of incorrect* B*3,0 results*

	table 1a: $\|E\| \geqslant 1.7$		table 1b: extended	
Range of $(AB + C)$	N(tot)	N(inv)	N(tot)	N(inv)
+200 . . . +70	43	1	42	0
+70 . . . +40	59	1	68	1
+40 . . . +10	162	6	107	8
+10 . . . 0	96	5	37	3
0 . . . −20	84	68*	58	38*
−20 . . . −60	16	5**	20	4**

N(inv) is the number of interactions contradicting eq. (2) (as observed after the structure determination).

* In this interval, most of the negative $(AB + C)$-values correspond to positive triple products.

** In this interval most of the negative $(AB + C)$-values correspond to negative triple products (invalid Σ_2-relations).

B*3,0 results.* The results were tabulated according to decreasing values of $(AB + C)$. Eleven letter symbols were assigned to the reflections with $|E| \geqslant 1.7$ occurring most often in the top half of the table. The sign correlation procedure (Beurskens, 1963) was used to generate symbolic signs from the $B3,0$ results. We define the following sets of reflections:

$\mathbf{h}_1$ are the eleven initial choices
$\mathbf{h}_2$ are reflections $\mathbf{h}_1 + \mathbf{h}'_1$
$\mathbf{h}_3$ are reflections $\mathbf{h}_1 + \mathbf{h}_2$ and $\mathbf{h}_2 + \mathbf{h}'_2$.

To find the reflections $\mathbf{h}_2$ and $\mathbf{h}_3$, only the top half of the $B3,0$-table was used. Several of the reflections were found at least three times with the same letter symbol. It is highly improbable that consistent results are obtained from incorrect $B3,0$ results, and therefore the multiple sign indications were assumed to be correct and these reflections added to the set $\mathbf{h}_1$. Thereafter, new reflections $\mathbf{h}_2$ and $\mathbf{h}_3$ could be calculated, and new reflections with multiple sign indications were added to the set $\mathbf{h}_1$, etc. Relations between the letter symbols were accepted as being correct where they were found at least five times, without inconsistencies. Finally the origin was fixed and two unknown letters remained to express the signs of 120 reflections $\mathbf{h}_1$.

The set $\mathbf{h}_1$ then was used as input data to a Σ_2-sign generation procedure, leading to the determination of a total of 514 signs. The weaker sigma-1 type $B3,0$ results were used at the end of the procedure to eliminate one of the letter symbols. The most probable of the two remaining solutions clearly revealed all of the carbon atoms of the molecule.

Structure refinement. The coordinates of the carbon atoms were refined, first with isotropic and then with anisotropic thermal factors. The fourteen hydrogen atoms attached to benzene carbon atoms were placed at calculated positions (C–H distance is 1.084 Å) and were included as constants in the refinement. The four hydrogen atoms attached to the carbon atoms of the cyclobutane ring were located on a difference-Fourier map and were included as variables in the last stage of the refinement. The temperature factors of all hydrogen atoms were fixed at a value of 4.0 Å^2. The full-matrix least-squares refinement was carried out on the function $\Sigma w(|F_o| - |F_c|)^2$, where $1/w = \sigma_c^2 + (0.05|F_o|)^2$ with σ_c the standard deviation calculated from counting statistics. The final R-value was 0.06 for all non-zero reflections. Structure factor calculations for 1519 reflections, including 549 unobserved reflections, gave $R = 0.12$. The atomic scattering factors used are those listed in the International Tables for X-ray Crystallography Vol. III. The fractional coordinates and thermal parameters, with standard deviations, of the carbon atoms are listed in table 2. Table 3 contains the coordinates of the hydrogen atoms. The structure is illustrated in figures 1 and 2. The structure factors are listed in table 4.

Discussion

The distribution of correct and incorrect $B3,0$ results is shown in table 1. About 12% of all calculated interactions do not satisfy the Σ_2-relationship; these invalid interactions may give rise to false solutions in standard direct methods techniques. By using only the top half of the $B3,0$ results we avoided about 75%* of the inconsistencies, and by using a careful sign generation procedure the probablity of making mistakes is greatly reduced. It was found *à postériori* that the four strongest reflections, all entering into many Σ_2-relations, are involved in eight non-valid Σ_2-interactions with

* correction june 1976: read 85%

Table 2. *Fractional coordinates and thermal parameters (×10⁴), with esd in parentheses. The anisotropic thermal parameters are in the form:*

$$\exp[-(h^2\beta_{11}+k^2\beta_{22}+l^2\beta_{33}+2hk\beta_{12}+2hl\beta_{13}+2kl\beta_{23})].$$

The key to atomic numbering is given in figure 1

Atom	x	y	z	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
C(1)	0.3882(2)	0.0096(6)	0.2104(4)	17(1)	133(9)	57(5)	9(3)	12(2)	11(6)
C(2)	0.4416(2)	−0.0199(6)	0.2712(4)	18(1)	97(8)	61(5)	0(3)	16(2)	2(6)
C(3)	0.4545(2)	−0.0503(6)	0.1584(4)	15(1)	115(9)	61(5)	−1(3)	13(2)	−11(6)
C(4)	0.3991(2)	−0.0771(7)	0.1092(5)	18(1)	128(10)	79(5)	5(3)	13(2)	14(6)
C(5)	0.4711(3)	0.0893(7)	0.3475(4)	20(1)	92(9)	57(5)	−2(3)	19(2)	0(6)
C(6)	0.5224(3)	0.0746(6)	0.3836(4)	18(1)	84(9)	53(5)	−2(3)	15(2)	−2(6)
C(7)	0.5486(2)	0.1763(7)	0.4501(5)	23(1)	111(9)	58(5)	−8(3)	11(2)	0(6)
C(8)	0.5250(3)	0.2929(7)	0.4830(5)	29(2)	102(11)	91(6)	−10(3)	24(3)	−22(6)
C(9)	0.4741(3)	0.3047(7)	0.4481(6)	30(2)	120(11)	87(6)	7(4)	32(3)	−4(7)
C(10)	0.4474(2)	0.2037(7)	0.3816(5)	24(1)	97(9)	66(5)	2(3)	19(2)	−2(6)
C(11)	0.3456(2)	−0.0286(7)	0.2588(5)	17(1)	118(10)	77(6)	0(3)	10(2)	−2(6)
C(12)	0.3450(2)	−0.1519(8)	0.3176(5)	16(1)	160(11)	96(6)	4(3)	14(2)	4(7)
C(13)	0.3058(3)	−0.1848(8)	0.3640(5)	21(1)	215(14)	91(6)	−3(4)	17(3)	16(7)
C(14)	0.2676(3)	−0.0948(11)	0.3538(7)	21(2)	239(16)	125(8)	−1(4)	21(3)	17(10)
C(15)	0.2673(3)	0.0275(10)	0.2979(7)	16(1)	277(18)	150(8)	21(4)	15(3)	−2(10)
C(16)	0.3064(3)	0.0624(7)	0.2466(6)	17(1)	175(12)	146(7)	17(4)	17(3)	18(8)
C(17)	0.3837(2)	−0.2326(7)	0.1018(5)	17(1)	135(12)	75(6)	−6(3)	17(2)	−22(7)
C(18)	0.4126(2)	−0.3367(8)	0.1643(6)	22(1)	102(10)	109(7)	−7(3)	18(2)	−10(7)
C(19)	0.3983(3)	−0.4778(8)	0.1551(6)	23(2)	137(12)	133(8)	−1(4)	26(3)	−13(8)
C(20)	0.3552(4)	−0.5148(9)	0.0821(9)	27(2)	203(16)	179(11)	−24(5)	28(4)	−54(11)
C(21)	0.3253(3)	−0.4117(12)	0.0205(7)	21(2)	265(18)	133(9)	−26(5)	7(3)	−38(11)
C(22)	0.3410(3)	−0.2706(10)	0.0303(6)	19(1)	228(15)	87(7)	−13(4)	6(3)	−10(8)

Table 3. *Hydrogen atom coordinates.*
H(1)–H(4): refined positions.
H(7)–H(22): calculated positions.
The numbering of the hydrogens refers to the parent carbon atoms.

Atom	*x*	*y*	*z*
H(1)	0.384(2)	0.128(7)	0.190(5)
H(2)	0.442(2)	−0.123(7)	0.310(5)
H(3)	0.479(2)	−0.131(7)	0.143(5)
H(4)	0.386(2)	−0.017(7)	0.041(5)
H(7)	0.588	0.166	0.478
H(8)	0.546	0.373	0.535
H(9)	0.456	0.394	0.474
H(10)	0.408	0.213	0.356
H(12)	0.376	−0.224	0.328
H(13)	0.306	−0.283	0.408
H(14)	0.237	−0.120	0.390
H(15)	0.237	0.100	0.292
H(16)	0.305	0.158	0.199
H(18)	0.446	−0.307	0.221
H(19)	0.421	−0.557	0.205
H(20)	0.345	−0.625	0.073
H(21)	0.291	−0.440	−0.034
H(22)	0.319	−0.191	−0.020

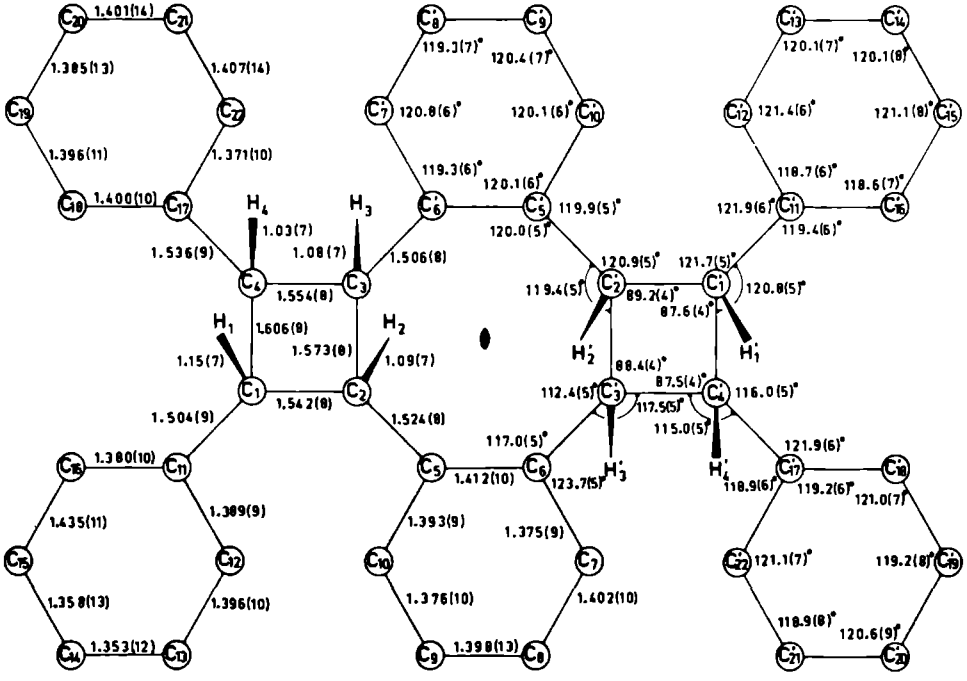


Fig. 1. Bond distances and angles, with esd in parentheses.

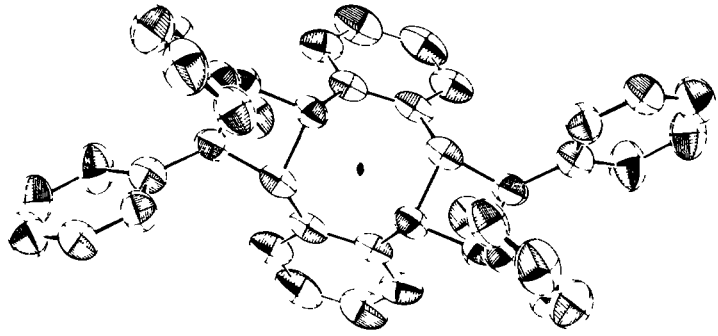


Fig. 2. Thermal ellipsoid plot of the molecule, as seen along the twofold rotation axis.

reflections having $|E| \geqslant 1.7$. Only one of these interactions is present in the top half of the table of $B3,0$ results.

The $B2,0$ formula appears to be very useful in saving computer time. It is seen from comparing tables 1a and 1b that the introduction of reflections with smaller $|E|$-values but selected for larger $|E'|$-values, does not reduce the reliability of the $B3,0$ results.

In our opinion, the procedure used in this structure investigation can never compete with standard techniques as far as simplicity and computer time is concerned. It may be useful only in cases where routine techniques have failed. As the main feature of the present procedure seems to be the avoidance of incorrect Σ_2-interactions, it should be possible to apply the same principles to small acentric structures.

The molecular structure is found to be in agreement with formula (II). So the argument that no cyclobutane moiety can be present in the dimer because of lack of stilbene formation on pyrolysis is wrong, obviously. This conclusion was reached recently in other investigations. Laarhoven & Cuppen (1972) prepared some compounds in which thermolysis of a 1,2-diphenylcyclobutane moiety apparently proceeds in a regiospecific way: *cis*-, *cis*-, *cis*-1,2,2a,10b-tetrahydro-1,2-diphenylcyclobuta[1]phenantrene gives stilbene on thermal decomposition, but the *trans*-, *trans*-, *trans*-isomer does not.

Meinwald & Young (1971) reported that a dimer of 1,8-distyrylnaphthalene, containing two *cis*-, *trans*-, *cis*-1,2-diphenylcyclobutane parts gives stilbene on heating, but a dimer of 2,3-distyrylnaphthalene with identically substituted cyclobutane rings does not (Ottenheijm, 1973). From these facts, it is not quite possible to predict whether or not a 1,2-diphenylcyclobutane moiety in a molecule will produce stilbene on pyrolysis. The present compound can be formed by a head-to-head dimerization of *trans*-, *trans*-*o*-distyrylbenzene, possibly *via* an eximer.

The geometry of the molecule is shown in figure 1. The molecule possesses a twofold rotation axis through the dodecadiene ring. The dihedral angle

between the benzo-groups is 61.5°. The cyclobutane rings are *cis*-, *trans*-, *cis*-substituted. The molecular symmetry (C_2) is far from $mm2(C_{2v})$ because of the non-planarity of the cyclobutane rings. The deviations of the atoms C(1) through C(4) from the best plane through the cyclobutane ring are 0.139, −0.142, 0.141 and −0.138 Å, respectively, all ±0.006 Å. The two dihedral angles formed by the pairs of planes through three carbon atoms having a diagonal in common are both 152°. This angle is in the range of 149-155°, reported for many puckered cyclobutane derivatives (Andreetti et al, 1973, and many references therein; Margulis, 1965; Adman & Margulis, 1967).

In the literature, there are two more structure determinations of *cis*-, *trans*-, *cis*-substituted cyclobutane containing compounds described: tetracyanocyclobutane (Greenberg & Post, 1968) and cyclobutanetetracarboxylic acid (Margulis, 1971). In both of them, however, the cyclobutane rings are planar.

As is usually found in tetra-substituted cyclobutane rings, the *cis*-substituted carbon atoms (1.573(8), 1.606(8) Å) are significantly longer than a single C–C bond of 1.537 Å; the distances between *trans*-substituted carbon atoms (1.542 and 1.554 Å) are only slightly longer.

The packing of the molecules in the unit cell is in agreement with van der Waals distances; no unusual intermolecular contacts occur.

Acknowledgements

The authors wish to express their thanks to Mr W. P. Bosman, Dr J. H. Noordik, Dr P. Benci and Mr J. M. M. Smits for valuable assistance in the X-ray work. The investigations were supported in part by 'FOMRE' with financial aid from the Netherlands Organization for Advancement of Pure Research (ZWO).

References

Adman, E. & Margulis, T. N. (1967) *Chem. Comm.* 641.
Andreetti, G. D., Bocelli, G. & Sgarabotto, P. (1973) *Cryst. Struct. Comm.* **2**, 115.
Beurskens, P. T. (1963) Technical Report on Sign Correlation by the Sayre Equation. The Crystallography Laboratory, Univ. of Pittsburgh, Pittsburgh, Pennsylvania.
Duax, W. L., Weeks, C. M. & Hauptman, H. (1972) *Acta Cryst.* **B28**, 1857.
Greenberg, B. & Post, B. (1968) *Acta Cryst.* **B24**, 918.
Hauptman, H & Karle, J. (1957) *Acta Cryst.* **10**, 267.
Hauptman, H. & Karle, J. (1958) *Acta Cryst.* **11**, 149.
Kanters, J. A., Kroon, J., Beurskens, P. T. & Vliegenthart, J. A. (1966) *Acta Cryst.* **21**, 990.
Karle, J. & Hauptman, H. (1957) *Acta Cryst.* **10**, 515.
Karle, J. & Hauptman, H. (1958) *Acta Cryst.* **11**, 264.
Karle, J. & Hauptman, H. (1959) *Acta Cryst.* **12**, 404.

Laarhoven, W. H., Cuppen, Th. J. H. M. & Nivard, R. J. F. (1970) *Tetrahderon* **26**, 1069.
Laarhoven, W. H. & Cuppen, Th. J. H. M. (1972) *J. Chem. Soc.* Perkin I, 2074.
Margulis, T. N. (1965) *Acta Cryst.* **19**, 857.
Margulis, T. N. (1971) *J. Am. Chem. Soc.* **93**, 2193.
Meinwald, J. & Young, J. W. (1971) *J. Am. Chem. Soc.* **93**, 725.
Muller, E., Sauerbier, M. & Heiss, J. (1966) *Tetrahedron Letters*, 2473.
Muller, E., Meier, H. & Sauerbier, M. (1970) *Chem. Ber.* **103**, 1356.
Ottenheijm, H. C. J. (1973) Personal communication of unpublished results.

CHAPTER 9

CRYSTAL STRUCTURE OF 'DIOX'

The paper published in Cryst. Struct. Comm.(1974) 3, 703 - 706 will be reproduced here.

2-(DI-*p*.ANISYLMETHYL)-1,3-DIOXOLANE, $C_{18}H_{20}O_4$

Th.E.M. van den Hark, H.M. Hendriks and Paul T. Beurskens

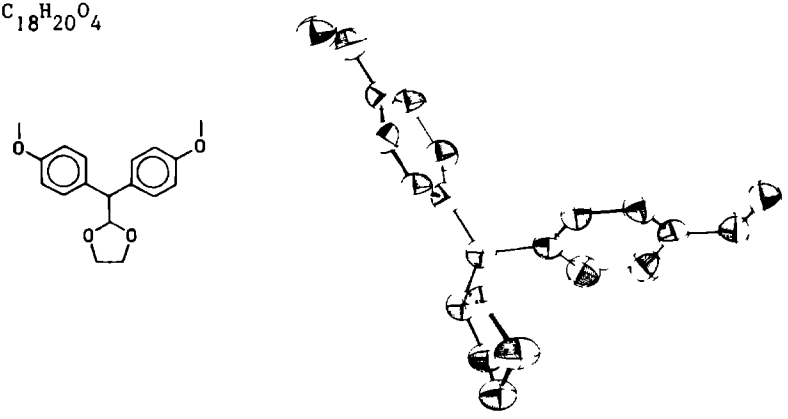
Crystallography Laboratory, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands.

Preliminary information. On request of Dr. Muizebelt we have solved the crystal structure of a reaction product, m.p. 107 - 108°, that was incorrectly 'identified' as 2,3-di-*p*-anisyldioxan (Muizebelt, 1971); the present structure analysis showed the compound to be 2-(di-*p*.anisylmethyl)-1,3-dioxolane.

Crystal data. Orthorhombic crystals. From diffractometer readings: a = 15.815(5), b = 16.90(1), c = 5.789(2) Å. Space group $P2_12_12_1$. Z = 4. D = 1.288 g/cm^3.

Intensity data, structure determination and refinement. Intensity data were collected with Ni-filtered Cu-radiation on a computer controlled NONIUS three-circle diffractometer ($\theta - 2\theta$ scan). Of the 1551 attainable symmetry-independent reflections within the limit $\theta<65°$, 1410 reflections were observed above background ($I>3\sigma_c(I)$, $\sigma_c(I)$ based on counting statistics). The structure was solved by direct methods: in the first stage of the analysis the B3,0 formula was applied to obtain phases of 76 projection reflections. This formula, derived by Karle and Hauptman (1958), was applied as described before(Kanters et al., 1966; van den Hark et al., 1974; Beurskens, 1974). In the second stage 10 general hkl reflections were given letter-symbols and the phase correlation procedure (Beurskens, 1964, 1974) was used

$C_{18}H_{20}O_4$



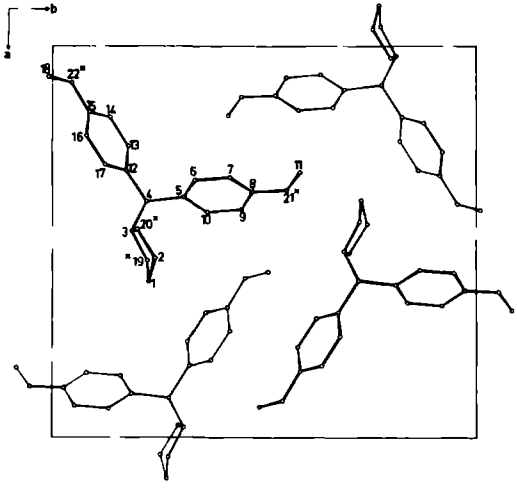
Fractional atomic coordinates.

	x	y	z		x	y	z
C(1)	0.6039(2)	0.2285(3)	0.9346(7)	C(12)	0.3180(2)	0.1752(2)	0.7123(5)
C(2)	0.5433(3)	0.2435(3)	1.1271(8)	C(13)	0.2565(2)	0.1801(2)	0.5415(6)
C(3)	0.4742(2)	0.1884(2)	0.8140(6)	C(14)	0.1819(2)	0.1367(2)	0.5546(7)
C(4)	0.3992(2)	0.2233(2)	0.6817(6)	C(15)	0.1689(2)	0.0860(2)	0.7377(7)
C(5)	0.3880(2)	0.3109(2)	0.7326(6)	C(16)	0.2293(2)	0.0800(2)	0.9100(7)
C(6)	0.3429(2)	0.3377(2)	0.9209(6)	C(17)	0.3034(2)	0.1242(2)	0.8939(7)
C(7)	0.3360(2)	0.4191(2)	0.9714(6)	C(18)	0.0786(3)	-0.0099(3)	0.9125(9)
C(8)	0.3735(2)	0.4731(2)	0.8237(6)	O(19)	0.5497(1)	0.2253(2)	0.7390(4)
C(9)	0.4188(2)	0.4474(2)	0.6346(7)	O(20)	0.4690(2)	0.2015(2)	1.0582(4)
C(10)	0.4259(2)	0.3666(2)	0.5892(7)	O(21)	0.3699(2)	0.5531(2)	0.8504(6)
C(11)	0.3232(3)	0.5846(3)	1.0355(9)	O(22)	0.0943(2)	0.0446(2)	0.7313(6)

Bond distances and angles

C(1) - C(2)	1.492(8)Å	C(5) - C(6)	1.380(7)Å	C(12) - C(13)	1.390(7)Å
C(1) - O(19)	1.421(7)	C(5) - C(10)	1.391(7)	C(12) - C(17)	1.380(7)
C(2) - O(20)	1.430(7)	C(6) - C(7)	1.411(7)	C(13) - C(14)	1.391(7)
C(3) - C(4)	1.531(7)	C(7) - C(8)	1.384(7)	C(14) - C(15)	1.379(7)
C(3) - O(19)	1.416(6)	C(8) - C(9)	1.378(7)	C(15) - C(16)	1.385(7)
C(3) - O(20)	1.434(6)	C(8) - O(21)	1.364(6)	C(15) - O(22)	1.372(6)
C(4) - C(5)	1.519(6)	C(9) - C(10)	1.396(7)	C(16) - C(17)	1.393(7)
C(4) - C(12)	1.529(6)	C(11)- O(21)	1.406(8)	C(18) - O(22)	1.417(8)

Atoms 19 - 22 (marked x) are oxygen atoms



Bond distances and angles (cont.)

C(1) - C(2) - O(20)	103.6(4)°	C(6) - C(7) - C(8)	118.8(4)°
C(1) - O(19)- C(3)	106.3	C(7) - C(8) - C(9)	120.4
C(2) - O(20)- C(3)	107.8	C(7) - C(8) - O(21)	124.5
C(2) - C(1) - O(19)	102.4	C(8) - C(9) - C(10)	119.9
C(3) - C(4) - C(5)	111.6	C(8) - O(21)- C(11)	118.9
C(3) - C(4) - C(12)	112.8	C(9) - C(8) - O(21)	115.1
C(4) - C(3) - O(19)	109.3	C(12)- C(13)- C(14)	121.6
C(4) - C(3) - O(20)	112.9	C(12)- C(17)- C(16)	121.9
C(4) - C(5) - C(6)	122.3	C(13)- C(12)- C(17)	117.5
C(4) - C(5) - C(10)	119.6	C(13)- C(14)- C(15)	119.8
C(4) - C(12)- C(13)	118.3	C(14)- C(15)- C(16)	119.8
C(4) - C(12)- C(17)	124.2	C(14)- C(15)- O(22)	115.1
C(5) - C(4) - C(12)	113.5	C(15)- C(16)- C(17)	119.5
C(5) - C(6) - C(7)	121.6	C(15)- O(22)- C(18)	117.5
C(5) - C(10)- C(9)	121.1	C(16)- C(15)- O(22)	125.1
C(6) - C(5) - C(10)	118.1	O(19)- C(3) - O(20)	106.4

to obtain the phases of 56 general reflections. By elimination or rejection of letter-symbols a twofold solution was obtained. In the final stage of the analysis the tangent formula was used to obtain some 300 phases for each of the two solutions. The positions of all non-hydrogen atoms were easily found on one of the calculated Fourier maps. Atomic

coordinates and anisotropic thermal parameters of all non-hydrogen atoms were refined by means of a full-matrix least-squares program, minimizing the function $\Sigma\, w\, \{|F_o| - |F_c|\}^2$, with weigts $w = \{\sigma_c^2\,(F_o) + (0.05\,F_o)^2\}^{-1}$.
The phenyl hydrogen atoms were placed at calculated positions, the remaining hydrogens were found on a Fourier difference map.
The hydrogens were included for the structure factor calculations, with $B = 3.0\ \text{Å}^2$. Refinement was continued, with fixed hydrogen parameters, until convergence was achieved. The final conventional R-value is 0.051 for all 1410 observed reflections.

Acknowledgements. The investigations were supported in part by the Netherlands Organization for the Advancement of Pure Research (ZWO/FOMRE). We wish to thank Drs. J.H. Noordik and Th.W. Hummelink for their contributions to this research.

References.

Beurskens, P.T. (1964). Acta Cryst. 17, 462.

Beurskens, P.T. (1974). Direct Methods in Crystallography (Lecture Notes, p. E1) Erice, Sicily.

Hark, Th.E.M. van den, Beurskens, P.T. and Laarhoven, W.H. (1974). J. Cryst. Mol. Str. 4, in the press.

Kanters, J.A., Kroon, J., Beurskens, P.T. and Vliegenthart, J.A. (1966). Acta Cryst. 21, 990.

Karle, J. and Hauptman, H. (1958). Acta Cryst. 11, 264.

Muizebelt, W.J. (1971). Tetrahedron Letters 51, 4879.

Received: 28 August 1974

CHAPTER 10

THE SIGMA-3 AND PAIR RELATIONSHIPS

In chapter 6 we defined a structure invariant as a quantity which is independent of the choice of origin. Analogously, it can be said that the value of a structure invariant is not affected by an arbitrary shift of the position of the origin. When an expression for the structure factor is chosen, there are, in general, in addition to the structure invariants, more quantities which have the same value for every choice of origin permitted by the chosen expression for the structure factor. All quantities which are invariant under a shift from one (permitted) origin position to another one are called structure seminvariants. Thus, the value of a structure seminvariant is determined by the structure alone. The cosine of a structure seminvariant is said to be a cosine seminvariant. In general, the values of the cosine seminvariants are obtainable from the observed magnitudes of the structure factors.
Example: in the space group $P2_1$ the origin is usually chosen on one of the two-fold screw axes. The origin may be shifted from one screw axis to another one and also along the screw axes. Then all phases $\phi_{\underline{h}}$ of reflections $\underline{h}$ = (2h 0 21) are structure seminvariants.

A pair relationship is an expression for $\phi_{\underline{h}_1} + \phi_{\underline{h}_2}$. It is necessary that the reflections $\underline{h}_1$ and $\underline{h}_2$ are chosen in such a way that $\phi_{\underline{h}_1} + \phi_{\underline{h}_2}$ is a structure seminvariant. For the space group $P2_1$, for example, $\underline{h}_1$ and $\underline{h}_2$ are chosen such that $\underline{h}_1 + \underline{h}_2$ = (2h 0 21).

In contrast to the well known sigma-1 and sigma-2 formulae, the other sigma formulae are not often used. The sigma-3 formula was an especially important tool in the early days of direct methods (see Hauptman and Karle, 1953). It has been shown that the sigma-3 relationship is very useful in selecting the correct solution out of many solutions which are equally probable with respect to the sigma-2 relationship (Beurskens, 1965).

Recently, another pair relationship was published (Hauptman, 1972) which is very like the sigma-3 relationship, but the sigma-3 relationship and the new pair relationship give different results. We will restrict the discussion here to non-centric space groups with just one centric

projection, and the formulae given below are for space group $P2_1$ only. Define:

$$R_1 = \sum_k \alpha \, (|E_{h'k1'}|^2 - 1) \tag{10.1}$$

$$R_2 = (-1)^{k_1} \sum_k \alpha \, (|E_{h''k1''}|^2 - 1) \tag{10.2}$$

$$R_3 = \sum_k \alpha \, \{(|E_{h'k1'}|^2 - 1)(|E_{h''k+k_1 1''}|^2 - 1)\} \tag{10.3}$$

with $\alpha = (-1)^k$, and define further $2h' = h_1+h_2$, $21' = l_1+l_2$, $2h'' = h_1-h_2$ and $21'' = l_1-l_2$.

We define now the following sigma-3 interactions:

$$\phi_{h_1k_1l_1} + \phi_{h_2\bar{k}_1l_2} \simeq \phi(R_1) \tag{10.4a}$$

$$\phi_{h_1k_1l_1} + \phi_{h_2\bar{k}_1l_2} \simeq \phi(R_2) \tag{10.4b}$$

The 1953-sigma-3 relationship (Hauptman and Karle, 1953) actually is a sum over interactions (10.4) where the phase of one reflection $h_1k_1l_1$ is given as a summation over all possible $h_2k_1l_2$ reflections, of which the phase should be known. The formula (10.4), derived for centric space groups, will be interpreted as follows: large positive or negative results (for R_1 and R_2) are strong indications that the seminvariant $\phi_{h_1k_1l_1} + \phi_{h_2\bar{k}_1l_2}$ is approximately equal to 0^o or 180^o, respectively.

Note. In case the reflection 2h' 0 21' is strong, the result for the summation R_1 gives the probable phase for the reflection 2h' 0 21'. A sigma-3 interaction can then be interpreted as a combination of a sigma-1 relationship and a sigma-2 interaction:

$$\phi_{h_1k_1l_1} + \phi_{h_2\bar{k}_1l_2} \simeq \phi_{2h'\ 0\ 21'} \simeq \phi(R_1) \tag{10.5}$$

If, however, the reflection 2h' 0 21' is weak (or outside the measured reflection sphere) (10.5) is no longer valid; the sigma-3 interaction (10.4a) still is valid.

Eq. (10.3) is a simplified form of the 1972-pair relationship (Hauptman, 1972):

$$|E_{h_1k_1l_1} E_{h_2k_1l_2}| \cos(\phi_{h_1k_1l_1} + \phi_{h_2\bar{k}_1l_2}) \simeq \tfrac{1}{2} N R_3 n_3^{-1} \tag{10.6}$$

with n_3 the number of reflections contributing to the summation (10.3). It was shown by Hauptman that when the right hand side of eq. (10.6) gives a large positive or negative result, the seminvariant

$\phi_{h_1k_1l_1} + \phi_{h_2\bar{k}_1l_2}$ is approximately equal to 0^o or 180^o, respectively.

We have found that the results are far more reliable if, for any one pair, R_1 , R_2 and R_3 all are large and have the same sign. We want to calculate a weight, w_{123}, for each pair, giving its reliability. Pairs with strong and consistent indications R_1 , R_2 and R_3 must obtain the largest weights.

Analogous to n_3, define n_1 and n_2 as the number of reflections which contribute to the summations (10.1) and (10.2). Further, we want to define a weight for the indication R_1. First of all, the weight should increase as the average contribution of the reflections to the summation (10.1) increases (also multiplication of R_1 with n_1^{-1}). Secondly, the reliability of the indication R_1 increases as the number of reflections contributing to the summation (10.1) increases (we prefer multiplication by $n_1^{\frac{1}{2}}$). Therefore, we define:

$$w_1 = R_1 \cdot n_1^{-\frac{1}{2}} \tag{10.7}$$

and analogously for the indications R_2 and R_3:

$$w_2 = R_2 \cdot n_2^{-\frac{1}{2}} \tag{10.8}$$

$$w_3 = R_3 \cdot n_3^{-\frac{1}{2}} \tag{10.9}$$

Note that the absolute value of w_3 has approximately the same magnitude as the 'significance level s' defined by Hauptman (1972).

First we combine the two sigma-3 indications (10.1) and (10.2) giving w_{12}:

$$w_{12} = 0.5\ S(w_1+w_2)\ \{S(w_1)|w_1|^{\frac{1}{2}} + S(w_2)|w_2|^{\frac{1}{2}}\}^2 \tag{10.10}$$

We will call w_{12} the 's.m.r.' (≡ square mean root) of w_1 and w_2. Some results of (10.10) are shown in table 10.1. It is clear that w_{12} has the largest value when w_1 and w_2 have comparable values and that w_{12} drops very rapidly when the difference between w_1 and w_2 increases. Combination of w_{12} and w_3 is done after scaling the w_{12} values such that the strongest results for w_{12} and for w_3 become comparable {the procedure that is used is given in the description of the computer program SIG3 (appendix A_1)}. The combination of w_{12} and w_3, giving w_{123}, is also done with the 's.m.r.' formula (10.10).

Table 10.1 Some numerical examples of the 's.m.r.' formula (10.10).

w_1	w_2	w_{12}
100	100	200
150	50	187
200	0	100
250	-50	38
300	-100	27

for all entries: $w_1 + w_2 = 200$

We will give some numerical results for two test structures, both in the space group $P2_1$, which were taken from literature:

a. Guanosine dihydrate, $C_{10}N_5O_5H_{13} \cdot 2H_2O$, Z = 4 (Thewalt et al., 1970).
   Calculations: I. 901 centric pairs for 71 h01 reflections with $|E|>1.0$
   II. 931 non-centric pairs for 273 reflections with $|E|>1.5$

b. Tetraphenyl ethene, $C_{26}H_{20}$, Z = 2 (Hoekstra, 1974).
   Calculations: I. 2019 centric pairs for 126 h01 reflections with $|E|>1.0$
   II. 1852 non-centric pairs for 420 reflections with $|E|>1.5$

In tables 10.2 and 10.3, the results are given for the top 50 and the top 100 pairs, respectively, for each structure. Non-centric pairs are considered incorrect if an error of more than 45° is made.

Table 10.2 Guanosine dihydrate. Number of incorrect pairs.

	50 top results	100 top results
for $\|w_{12}\|$	I : 3 II : 1	I : 16 II : 11
for $\|w_3\|$	I : 2 II : 1	I : 15 II : 7
for $\|w_{123}\|$	I : 0 II : 0	I : 11 II : 4

Table 10.3 Tetraphenyl ethene. Number of incorrect pairs.

	50 top results	100 top results
for $\|w_{12}\|$	I : 8 II : 18	I : 23 II : 37
for $\|w_3\|$	I : 6 II : 18	I : 17 II : 36
for $\|w_{123}\|$	I : 5 II : 16	I : 11 II : 35

As can be seen from tables 10.2 and 10.3, the combination of the results from the 1953-sigma-3 relationship and the 1972-pair relationship does make sense; in all cases, the reliability of the pairs using the combined weights, w_{123}, is better than for either w_{12} as for w_3. In general, there are only a limited number of reliable pairs: when going from the top 50 to the next 50 pairs, the reliability drops significantly and when going further down the list of pairs the reliability becomes even worse. For 'simple structures' such as centric structures or non-centric structures with three centric projections the results of the pair relationship are unimportant, but for 'difficult problems' such as structures in the space group $P2_1$ the results are very useful. It is important to realize that the information from the pair relationship is quite different from the information which is obtained from the sigma-2 relationship (e.g. the weak reflections play an important role in the pair relationship). We never just use the results of the pair relationship, but always in combination with calculated triple invariants.

References

Beurskens, P.T. (1965). Thesis, University of Utrecht.

Hauptman, H. (1972). Crystal Structure Determination. The Role of the Cosine Seminvariants. Plenum Press. New York.

Hauptman, H. and Karle, J. (1953). Solution of the Phase Problem, I. The Centrosymmetric Crystal. A.C.A. Monograph, No. 3.

Hoekstra, A. (1974). Thesis, University of Groningen.

Thewalt, U., Bugg, C.E. and Marsh, R.E. (1970). Acta Cryst. B26, 1089.

CHAPTER II

'HEPTA.NC'

11.1 Crystal structure of 'HEPTA.NC'

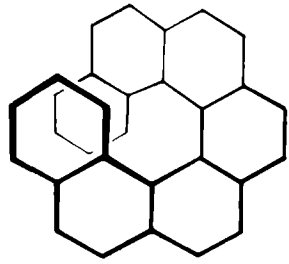
We will reproduce in this section the paper published in Cryst. Struct. Comm. (1976) 5, 000

HEPTAHELICENE, $C_{30}H_{18}$

Paul T. Beurskens, Gezina Beurskens and Th.E.M. van den Hark

Crystallography Laboratory, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands

Preliminary information. X-ray crystal structure determinations of helicenes and related compounds have recently been carried out: see Van den Hark et al. (1974) and Hummelink- Peters et al.(1975) and references given therein. The present compound was prepared by Laarhoven et al. (1970). This compound crystallizes in two modifications: we now report the structure in space group $P2_1$; the X-ray analysis of a structure in space group $P2_1/c$ has also been completed (Germain et al., 1975 and Van den Hark et al., 1976a).

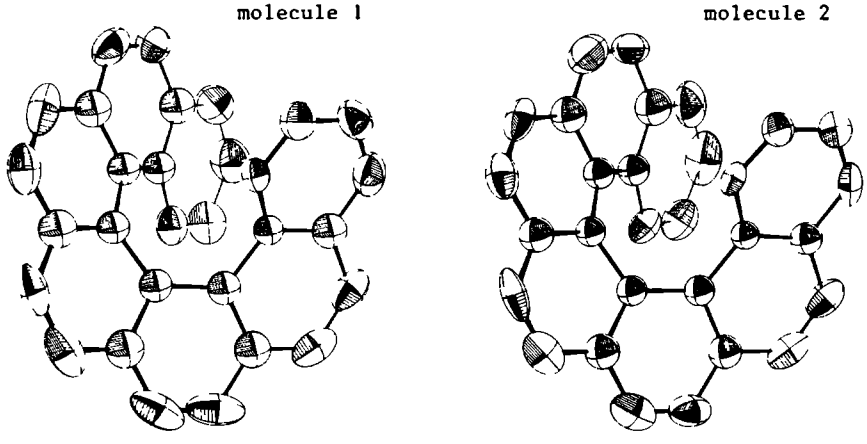


Crystal data. (From single crystal diffractometry, $\lambda MoK\alpha = 0.71069\text{Å}$). Monoclinic, $\underline{a} = 14.022(4)$, $\underline{b} = 15.094(4)$, $\underline{c} = 9.221(3)\text{Å}$, $\beta = 93.20(2)^{\circ}$ and $V = 1949\ \text{Å}^3$. Space group $P2_1$, $Z = 4$, $D_x = 1.288\text{g/cm}^3$, F.W. = 378.5.

Intensity data, structure determination and refinement. Intensity data of 3565 symmetry independent reflections up to $\sin\Theta/\lambda = 0.59\ \text{Å}^{-1}$ were collected on a NONIUS-CAD-3 diffractometer. 2022 reflections with

$I>3\sigma_c(I)$ ($\sigma_c(I)$ based on counting statistics) were used in the refinement.

The structure was solved by direct methods, using the B3,0 formula and the sigma-3 and pair relationships as described by Van den Hark et al. (1975) and Beurskens et al. (1975). The phases of the projection reflections h01 were expressed in terms of two letter symbols, leading to a fourfold solution. For the origin and enantiomorph fixation, strongly correlated sets of h21 reflections and h31 reflections were used. One of the four final solutions did reveal clearly recognisable molecules. The structure, however, could not be refined. We then used one heptahelicene molecule in space group P1 to phase the difference structure factors; direct methods were used to refine the phases, using the computer program DIRDIF.D (Van den Hark et al., 1976b). The result was a structural model in which the first molecule was shifted over 0.65 Å with respect to the two-fold screw axis. This model rapidly refined. The atomic coordinates, the anisotropic thermal parameters for the 18 outer carbon atoms of each molecule and the isotropic thermal parameters for the remaining carbon atoms were refined by full-matrix least-squares refinement techniques minimizing the function $\Sigma w\{|F_o| - k|F_c|\}^2$, with $w = \{\sigma_c(F_o)^2 + (0.05F_o)^2\}^{-1}$. Hydrogen atoms were placed on calculated positions with C-H = 1.084 Å, except for H(1), H(18) , H(31) and H(48), which were located on a subsequent difference Fourier synthesis. The hydrogen atoms were given an isotropic temperature factor $B = 5.0 Å^2$ in the structure factor calculations, but they were not refined. The

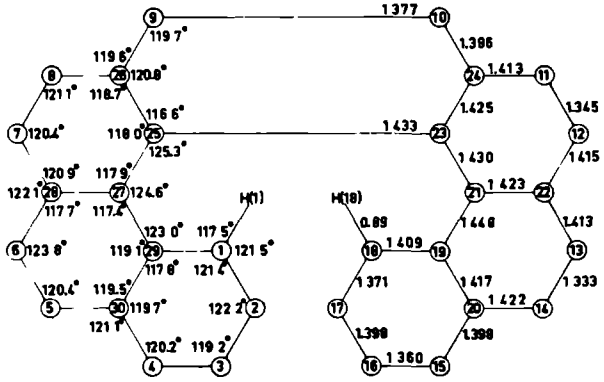


Atomic coordinates.

Heptahelicene molecule 1				Heptahelicene molecule 2			
Atom	x	y	z	Atom	x	y	z
e.s.d.:	0.0007	0.0007	0.0011	e.s.d.:	0.0007	0.0007	0.0011
C(1)	0.9796	0.4876	0.7919	C(31)	0.4305	0.5534	0.4930
C(2)	1.0413	0.5505	0.7430	C(32)	0.3406	0.5406	0.5429
C(3)	1.1006	0.5986	0.8440	C(33)	0.2626	0.5173	0.4476
C(4)	1.0922	0.5868	0.9858	C(34)	0.2779	0.5093	0.3038
C(5)	1.0103	0.5214	1.1893	C(35)	0.3761	0.5300	0.0939
C(6)	0.9404	0.4685	1.2365	C(36)	0.4601	0.5555	0.0455
C(7)	0.8091	0.3633	1.1927	C(37)	0.6291	0.6038	0.0813
C(8)	0.7557	0.3096	1.0991	C(38)	0.7068	0.6164	0.1674
C(9)	0.7238	0.2459	0.8607	C(39)	0.7892	0.6075	0.4043
C(10)	0.7479	0.2329	0.7173	C(40)	0.7920	0.5868	0.5477
C(11)	0.8635	0.2443	0.5338	C(41)	0.7199	0.5128	0.7528
C(12)	0.9501	0.2676	0.4937	C(42)	0.6518	0.4613	0.8057
C(13)	1.1168	0.3199	0.5546	C(43)	0.5184	0.3561	0.7650
C(14)	1.1879	0.3421	0.6522	C(44)	0.4616	0.3084	0.6823
C(15)	1.2480	0.3435	0.9055	C(45)	0.4148	0.2493	0.4396
C(16)	1.2397	0.3233	1.0483	C(46)	0.4312	0.2426	0.2932
C(17)	1.1528	0.2890	1.0939	C(47)	0.4998	0.2941	0.2373
C(18)	1.0763	0.2808	0.9952	C(48)	0.5496	0.3550	0.3182
C(19)	1.0826	0.3057	0.8494	C(49)	0.5348	0.3664	0.4672
C(20)	1.1736	0.3322	0.8027	C(50)	0.4689	0.3091	0.5293
C(21)	1.0024	0.2996	0.7430	C(51)	0.5866	0.4302	0.5593
C(22)	1.0243	0.2959	0.5937	C(52)	0.5869	0.4169	0.7117
C(23)	0.9059	0.2975	0.7816	C(53)	0.6435	0.5021	0.5093
C(24)	0.8374	0.2579	0.6778	C(54)	0.7188	0.5344	0.6024
C(25)	0.8688	0.3336	0.9105	C(55)	0.6284	0.5450	0.3686
C(26)	0.7834	0.2956	0.9564	C(56)	0.7094	0.5881	0.3131
C(27)	0.9132	0.4038	0.9953	C(57)	0.5399	0.5498	0.2851
C(28)	0.8875	0.4106	1.1422	C(58)	0.5419	0.5697	0.1348
C(29)	0.9749	0.4700	0.9406	C(59)	0.4461	0.5404	0.3439
C(30)	1.0281	0.5260	1.0385	C(60)	0.3670	0.5235	0.2461
H(1)	0.945	0.452	0.735	H(31)	0.483	0.570	0.541
H(18)	1.019	0.256	1.026	H(48)	0.590	0.386	0.268

final conventional R-value is 0.065 for the 2022 observed reflections.

Comments. The molecules possess a pseudo (local) two-fold symmetry axis; the deviations from true symmetry (with respect to bond distances) are in the order of the standard deviations. This is true also for the deviations between the two symmetry-independent molecules. Therefore corresponding bond distances and angles have been averaged (e.s.d. of the averages: 0.006Å and 0.5°).



Some individual contact distances for molecule 1, and similar distances for molecule 2, and the corresponding contact distances for tribenzoheptahelicene (Van den Hark et al., 1974) are given:

	mol. 1	mol. 2	tribenzo-der.
C(1) - C(21)	2.892(10)Å	2.910(10)Å	2.971Å
C(1) - C(23)	3.049	3.080	3.034
C(18) - C(27)	2.945	2.960	2.892
C(18) - C(25)	3.075	3.099	3.011
H(1) - C(23)	2.44	2.51	2.40
H(18) - C(25)	2.58	2.61	2.42

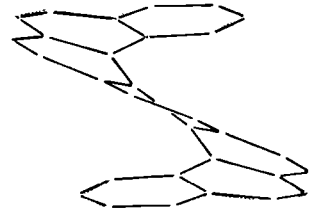
Torsional angles for the two molecules

	mol.1	mol.2
H(1) - C(1) - C(29) - C(27)	-4.8°	3.0°
C(1) - 29 - 27 - 25	16.8	20.7
29 - 27 - 25 - 23	26.1	21.8
27 - 25 - 23 - 21	23.4	26.3
25 - 23 - 21 - 19	26.0	26.7
23 - 21 - 19 - 18	21.9	14.2
21 - 19 - 18 - H(18)	3.0	-1.7

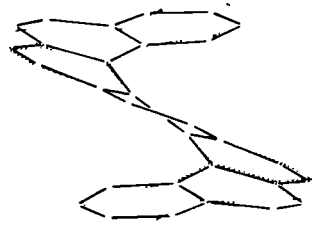
Interplanar angles for consecutive C_6-rings

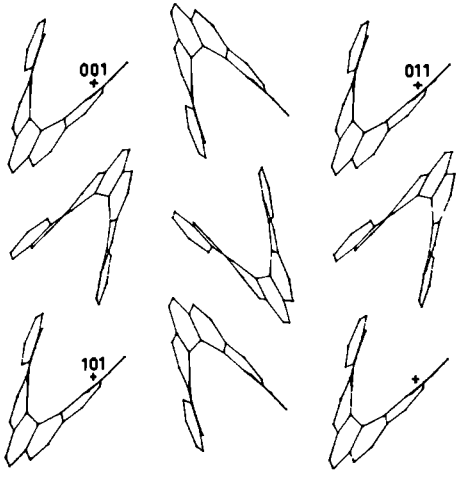
Planes	mol.1	mol.2
A,B	10.6°	12.2°
B,C	11.5	12.4
C,D	13.2	12.0
D,E	12.5	13.4
E,F	12.4	11.7
F,C	10.9	9.8
(A,G	30.7	33.8)

Each of the two independent molecules have been idealized with respect to their molecular two-fold axes; thereafter the two molecules are projected on top of each other to show the similarity of the molecular geometry.
(——— molecule 1, ... molecule 2)



Similarly: the average of the two molecules is projected on top of the heptahelicene-skeleton of tribenzoheptahelicene.
(——— average of 1 and 2, ... tribenzoderivative)





Molecular packing: projection onto the (a,b) plane.

Acknowledgements. The authors thank Dr. W.H. Laarhoven for suggesting this research, and for preparing the crystals, and Dr. J.H. Noordik, Mr. J.M.M. Smits, and Drs. W.P. Bosman for their contributions to this research. One of us (Th.E.M. van den Hark) acknowledges support of the Dutch Foundation for Pure Research, ZWO/FOMRE.

References.

Beurskens, P.T.and Van den Hark, Th.E.M. (1975). Acta Cryst. A31, S14.

Germain, G. and collaborators (1975). Private communication.

Hummelink-Peters, B.G.M.C., Van den Hark, Th.E.M., Noordik, J.H. and Beurskens, Paul T. (1975). Cryst. Struct. Comm. 4,281.

Laarhoven, W.H., Cuppen, Th.J.H.M. and Nivard, R.J.F. (1970). Tetrahedron 26, 4865.

Van den Hark, Th.E.M., Noordik, J.H. and Beurskens, Paul T. (1974). Cryst. Struct. Comm. 3, 443.

Van den Hark, Th.E.M. and Beurskens, P.T. (1975). Acta Cryst. A31, S17.

Van den Hark, Th.E.M. and Beurskens, Paul T. (1976a). Cryst. Struct. Comm. 5,

Van den Hark, Th.E.M., Prick, Peter and Beurskens, Paul T. (1976b). Submitted for publication.

11.2 Origin and enantiomorph specification for 'HEPTA.NC'

In space group $P2_1$, the origin is usually fixed with two projection reflections (i.e. h0l reflections) and one h11 reflection; these reflections must have large $|E|$-values and must be involved in many triples. The origin is fixed on one of the two-fold screw-axes by arbitrarily giving a phase of 0° or 180° to the two projection reflections; by assigning arbitrarily a phase of, say 0°, to the h11 reflection, the position of the origin on the two-fold screw axis is fixed unambiguously. The enantiomorph is fixed with another general reflection (i.e. a reflection hkl with $k \neq 0$) having a phase deviating significantly from 0° or 180°.

In the crystal structure determination of 'HEPTA.NC', only a few strong h11 reflections were available, but many strong h21 and h31 reflections. In this section we will describe the way in which the h21 and h31 reflections, in addition to the projection reflections, were used to fix the origin and enantiomorph. Instead of single h21 and h31 reflections, we used sets of reflections of which the phases were correlated. The correlation of the phases was done with calculated triple invariants (large $W_{\underline{hk}}$ values, see chapter 7) in which one projection reflection (with a known phase) was involved, in combination with strong pairs (i.e. with large $|w_{123}|$ values, see chapter 10). As an example, the results for the h21 reflections are given in fig. 11.1. All reflections in this figure are involved in at least 3 strong pair or triple relationships; all phases are either α or $\alpha+180^\circ$. Similarly, we had a set of seven h31 reflections with phases β or $\beta+180^\circ$.

Let us use the set of h31 reflections for origin fixation by defining $\beta = 0^\circ$. Then the origin is not fixed unambiguously; moving the origin an integer multiple of 1/3 along the two-fold screw axis does not affect the phases of the h31 reflections. (When in space group $P2_1$ the origin is shifted a distance Δy (fractional) along the two-fold screw axis, all phases of reflections hkl change by a value $\Delta\phi_{hkl}$ where $\Delta\phi_{hkl} = \Delta y \,.\, k \,.\, 360^\circ$.) As a consequence, all reflections hkl with $k \neq 3n$ will have a three-fold phase ambiguity. When an h21 reflection (see fig. 11.1) has a phase α with respect to a fixed origin on the two-fold screw axis, the phase of this reflection with respect to the origin at $\Delta y = 1/3$ and $\Delta y = 2/3$ will be $\alpha+240^\circ$ and $\alpha+120^\circ$, respectively.

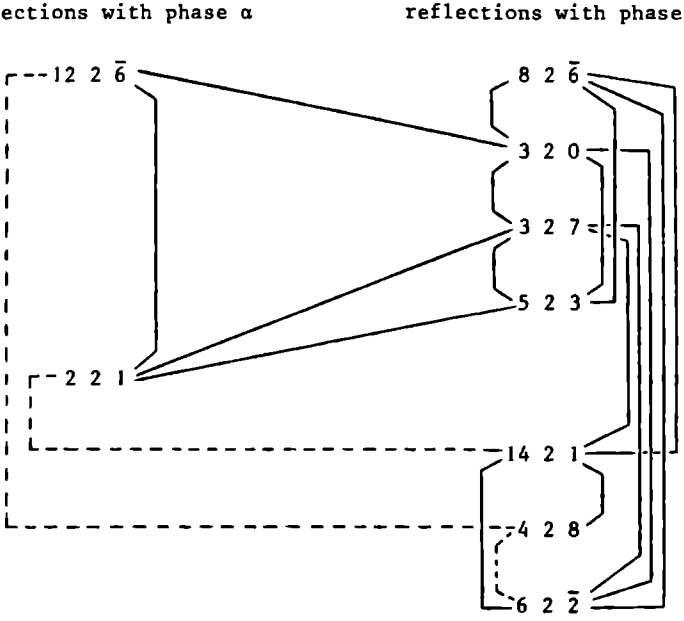


Fig. 11.1 The correlation of phases of h2l reflections.
——— *indicates a triple in which a projection reflection is involved.*
- - - - - - *indicates a pair relationship.*

The value of α is unknown, but it is clear that restricting the values of α to 1/3 of the phase circle completes the origin fixation. For example, α may have any value between -60° and $+60^{\circ}$.

The enantiomorph can be fixed by further restricting the possible values of α (except in case α happens to be 0°, 60° or -60°). When anomalous scattering is neglected, the two enantiomorphs of a structure have phases which differ in sign for all reflections. Therefore, restricting α to, for example, positive values does fix the enantiomorph. We decided to continue the phase determination with a value of 30° for α; then the true value of α deviates not more than 30° from this assignment. At this stage of the phase determination, we had 23 projection reflections and 16 general reflections with numerical phases or with phases expressed in terms of two letter symbols; both symbols were assigned to projection reflections and represent therefore phases of 0° or 180°.

11.3 The phasing process for 'HEPTA.NC'

Despite the fact that the phase determination was performed very carefully, none of the four solutions did reveal the correct crystal structure. In one of the Fourier maps, two heptahelicene molecules were found, which proved to have the correct orientation but an incorrect position in the unit cell. Both molecules appeared to be displaced over a vector perpendicular to the two-fold screw axis.

The difficulties encountered in solving the crystal structure of 'HEPTA.NC' can be explained from incorrect pairs and triples used for phase determination. It proved that many pairs and many triples relating phases of certain groups of reflections were incorrect. For example, almost all pairs

$$\phi_{hkl} + \phi_{h'\bar{k}l'} \quad \text{with } l<4 \text{ and } l'>4$$

and triples

$$\phi_{hkl} + \phi_{h'k'l'} + \phi_{-h-h' \; -k-k' \; -l-l'} \quad \text{with } l,l'<4 \text{ and } l+l'>4$$

proved to be incorrect, including the strongest ones.

The same type of problem arose in the crystal structure determination of the centric modification of heptahelicene, 'HEPTA.C' (chapter 17). In both structures ('HEPTA.NC' and 'HEPTA.C'), many C-C vectors are perpendicular to a symmetry-element, resulting in parallel symmetry-related C-C vectors. In section 17.2, the relation between the incorrect triples and the crystal structure will be discussed for 'HEPTA.C'.

APPENDIX A_1

THE NIJDIR SYSTEM

The NIJDIR system is a set of computer programs which is presently being developed at the Crystallography Laboratory of the University of Nijmegen, to solve crystal structures of equal atom compounds by direct methods. In addition to programs to solve structures routinely (such as the program SCOR) the system will provide facilities to tackle 'difficult' structures. In the preceding sections, we distinguished four categories of structures; we intend to include a set of program links to solve the phase problem fully automatically, for each of these categories.

A short description will be given of some of the programs.

ENTREX : the first program of the NIJDIR system. It prepares a NIJDIR reflection file from standard crystallographic reflection files.

TRIPEL : (sigma-2 listing) generates all triples that may be needed.

B30 : calculates triple invariants with the B3,0 formula for selected triples.

SIG3 : applies the sigma-3 and pair relationship.

SCHAAL : rescales the calculated AB_{hk} + C values and computes KAB_{hk} and W_{hk} values for all triples.

SCORB : application of the sign correlation procedure to the B3,0 results to solve centric structures or centric projections of non-centric structures.

PCOR : application of the phase correlation procedure.

TANGENT: refinement and extension of a phase set, obtained by other programs, with the tangent formula. In case not all letter symbols could be eliminated in preceding steps, the tangent formula is applied to each of the solutions.

These programs, except PCOR, are completed: presently, PCOR is executed with manual intervention. The automatic link between the programs will be subject to further research. Eventually, automatic Fourier and peak-searching programs will be included.

PART B

DIRECT METHODS APPLIED TO DIFFERENCE STRUCTURE FACTORS

CHAPTER 12

INTRODUCTION TO PART B OF THIS THESIS

12.1 Short description of the DIRDIF procedures

In the DIRDIF procedures direct methods are applied to solve structures where part of the structure is known (DIRDIF stands for: DIRect methods applied to DIFference structure factors). Usually the known part of the structure consists of heavy atoms in heavy atom structures. The tangent formula or sigma-2 relationship is applied to the difference structure factors in order to find, to verify or to modify the phases as well as the absolute values of the difference structure factors. The procedures are most useful when the known atoms do not fix the phases of the difference structure factors well, either because they are in special or pseudospecial positions, or because they are not sufficiently heavy relative to the other atoms. Automatic computer programs have been written in FORTRAN IV to execute the procedures. The programs are designed for routine structure analysis, and may be used without any previous experience of the applications of direct methods.

At present we distinguish the following procedures:

DIRDIF.A : the special case for centrosymmetric structures
DIRDIF.B : the general case for centrosymmetric structures
DIRDIF.C : the special case for non-centrosymmetric structures
DIRDIF.D : the general case for non-centrosymmetric structures

---

I am very thankful to Dr. R.O. Gould of the University of Edinburgh for his contribution to the development of the procedure DIRDIF.B and for much helpful discussion.

The special case is defined as the case where the known atoms are on special or pseudo-special positions, such that the origin (and/or enantiomorph) is not completely fixed by the known atoms. When the origin is not completely fixed, the known atoms do not contribute to all parity classes of reflections; the phases of these reflections are thus completely undetermined. The phases of one or more reflections can be fixed for complete origin specification. When the known atoms are in a centrosymmetric arrangement in a non-centrosymmetric structure, they do not fix the enantiomorph. After shifting the origin to a centre of symmetry all calculated phases are restricted to the values 0° or 180°. One reflection that is expected to have a phase of $\pm 90^{\circ}$ is used for enantiomorph specification. Non-centrosymmetric structures in which the known atoms fix the origin but not the enantiomorph also belong to the special case.

Solving the remaining unknown part of the structure in the special case with the conventional heavy atom technique is not straightforward. In a difference Fourier synthesis several structures are found mixed up; identifying the atoms belonging to one and the same structure may be difficult.

The general case is the case where the known atoms contribute to all classes of reflections, such that the origin (and enantiomorph) are completely fixed. Thus the phases of all reflections are in principle determined and it should be possible to find the complete structure by successive difference Fourier syntheses. However, direct methods can be used to minimize the errors in the phases of the difference Fourier coefficients, and to correct the amplitudes, before a synthesis is calculated.

Some examples

1. DIRDIF.A : space group $P\bar{1}$: two symmetry-independent positions 0,0,0 and 0,½,0 determine signs for reflections hkl with k=2n; one reflection hkl with k=2n+1 must be used to fix the origin completely.
2. DIRDIF.A : space group $P2_1/c$: one atom in the asymmetric unit of the unit-cell on the pseudo-special position x,¼,z determines signs for reflections hkl with l=2n; one reflection hkl with l=2n+1 must be used to fix the origin completely.
3. DIRDIF.B : space group $P\bar{1}$: one atom on 0,0,0.
4. DIRDIF.B : space group $P\bar{1}$: two symmetry-dependent atoms in general positions.

5. DIRDIF.C : space group P1 : one atom on 0,0,0; all calculated phases are 0°; select reflections that may have a phase of $\pm 90^{\circ}$ and accept one phase for enantiomorph specification.
6. DIRDIF.C : space group P1 : two independent atoms on $0,\frac{1}{4},0$ and $0,\frac{3}{4},0$; origin and enantiomorph specifications as in examples 1 and 5, respectively.
7. DIRDIF.D : space group P1 : two different or any non-symmetrical arrangement of three or more atoms.
8. DIRDIF.D : space group $P2_12_12_1$: one atom per asymmetric unit, on a general (not pseudo-special) position.

The procedures DIRDIF.A, DIRDIF.B and DIRDIF.D have been studied and applied to several crystal structure determinations. DIRDIF.C is in a preliminary stage of development.

The DIRDIF procedures were originally developed for heavy atom structures. When we were solving the crystal structures of 'HEPTA.C' and 'HEPTA.NC', we found in both cases a heptahelicene molecule in a correct orientation but at an incorrect position in the unit cell with respect to the symmetry-elements. The procedure DIRDIF.D was applied and proved to be a very convenient and effective tool for solving this type of problem.

Note: Application of direct methods to the structure factors of the complete structure will result in a redetermination of the heavy atom positions. For partially known equal atom structures, direct methods can also be applied with success to the structure factors of the complete structure (e.g. tangent recycling procedure; Karle, 1968). The advantage of our procedure will be discussed in chapter 18.

12.2 Procedure DIRDIF.A

The special case for centrosymmetric structures was studied by Beurskens and Noordik (1971) (see also Noordik, 1971). They developed a procedure to apply direct methods to centrosymmetric heavy atom structures with heavy atoms on special positions. A short description of that procedure will be given here.

Beurskens and Noordik distinguished two categories of reflections:

1. 'Strong' reflections. As the heavy atoms are on special positions, all of these reflections have an equal and maximal (all heavy atoms scatter

in phase) heavy atom contribution to a structure factor. On subtracting the heavy atom contribution from the observed structure factors, both the sign and the amplitude of the light atom contribution for these reflections are obtained (the heavy atoms are the dominant scatterers for these reflections).

2. 'Weak' reflections that do not have a contribution from the heavy atoms. The sign and the amplitude of the structure factor for these reflections are fully determined by the unknown light atom part of the structure only. Direct methods are used to find the signs of these reflections.

The procedure of Beurskens and Noordik consists of two steps. In the first step, scale and temperature factors are calculated. The following expression is used (Parthasarathy, 1966):

$$\langle I \rangle_{\underline{h}} = K_L \langle \overset{L}{\Sigma} f_L^2 \exp(-2B_L S^2) \rangle_{\underline{h}} + K_H \langle F_H^{o\,2} \exp(-2B_H S^2) \rangle_{\underline{h}} \qquad (12.1)$$

where $I = (K|F_{obs}|^2)$ is the observed intensity on a relative scale, $K = K_H = K_L$ is the scale factor, $\overset{L}{\Sigma}$ denotes a summation over all light atoms in the unit cell, F_H^o is the heavy atom contribution to the structure factor calculated with a temperature factor of zero, and B_H and B_L are the overall temperature factors of the heavy and light atoms, respectively. The average is taken over all reflections $\underline{h}$ within a given $\sin\Theta$ interval. $S = \sin\Theta/\lambda$.

For the weak reflections the second term in (12.1) vanishes and a Wilson plot for these reflections will give K_L and B_L. On substituting these results in eq. (12.1) a Wilson plot for the strong reflections will give values for B_H and K_H. Thereafter, normalized structure factors are calculated for the light atom part of the structure. For strong reflections the difference structure factor F_L is calculated from: $F_L = S(F_H) \cdot |F_{obs}| - F_H$, with $S(F_H)$ the sign of the heavy atom contribution to the structure factor; for the weak reflections $|F_L| = |F_{obs}|$.

In the second step the sign correlation procedure is used to sign the weak reflections with large $|E|$-values. The sign determination is started from the large group of signs of strong reflections and one or two signs of weak reflections, which may be chosen to fix the origin completely. After the sign determination, a Fourier map, calculated with the signed difference structure factors, will reveal the positions of the light atoms.

The procedure of Beurskens and Noordik has been the basis for the procedure DIRDIF.A but, as a consequence of our investigations on DIRDIF.B, DIRDIF.A can now also be applied to structures of a more general complexity. This new procedure, which will not be described in full detail in this thesis (see van den Hark and Beurskens, 1975), can also handle structures in which

a. the heavy atoms are on special positions, but are not sufficiently dominating to fix almost all signs of the difference structure factors of the strong reflections.
b. the heavy atoms are on pseudo-special positions (see example 2, section 12.1). In this case the heavy atom contribution to the structure factor for the strong reflections may have any value between zero and its maximum possible value. Now the strong reflections must be treated in the same way as in the procedure DIRDIF.B (see chapter 14).
   Note: the average intensity of the strong reflections is greater than the average intensity of the weak reflections, but individual reflections may have any intensity.

12.3 Design of part B

In chapter 13 an example will be given of the application of the procedure DIRDIF.A to the crystal strycture determination of 'KRh'. Chapter 14 deals with a description of the procedure DIRDIF.B (see also Gould et al., 1975). An example of the application of DIRDIF.B to the crystal structure determination of 'RUBIF' is given in chapter 15. The procedure DIRDIF.D is presented in chapter 16. The crystal structure determination of 'HEPTA.C' is described in chapter 17. In the last chapter of part B the application of the DIRDIF.D procedure to partially known equal atom structures is discussed.

12.4 References

Beurskens, P.T. and Noordik, J.H. (1971). Acta Cryst. A27, 187.

Gould, R.O., van den Hark, Th.E.M. and Beurskens, P.T. (1975). Acta Cryst. A31, S17.

van den Hark, Th.E.M. and Beurskens, P.T. (1975). Report 1975/1, Crystallography Laboratory, University of Nijmegen.

Karle, J. (1968). Acta Cryst. B24, 182.

Noordik, J.H. (1971). Thesis, University of Nijmegen.

Parthasarathy, S. (1966). Z. Kristallogr. 123, 27.

CHAPTER 13

CRYSTAL STRUCTURE OF 'KRh'

In this chapter the paper on the 'KRh' structure is reproduced, as an example of the application of the procedure DIRDIF.A. For brevity, table 5, thermal parameters and the structure factor table have been omitted.

Accepted for publication in Journal of the Chemical Society
(Dalton Transactions)

The Crystal and Molecular Structure of Potassium Bis(dimethylphenylphosphine)bis(dithiocarbonato)rhodate(III) Trihydrate

By Robert O. Gould and Alistair M. Gunn, Chemistry Department, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, and Thijs E.M. van den Hark, Crystallography Laboratory, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands.

The structure of the title compound, $K^{+}\{Rh(S_2CO)_2(PMe_2Ph)_2\}^{-}.3H_2O$ has been determined. Crystals are monoclinic, space group $P2_1/c$, with $\underline{a}$ = 18.798, $\underline{b}$ = 6.516, $\underline{c}$ = 23.924Å, β = 113.91°. There are two independent anions per unit cell with crystallographic $\underline{C}_i$ ($\bar{1}$) symmetry. They are essentially identical, and the planar dithiocarbonate ligands have Rh-S = 2.37, C-S = 1.725, C-O = 1.25Å; S-Rh-S = 73.5, S-C-S = 111 and S-C-O = 125°.

In a recent publication [1], the reaction between $\underline{mer}$-$\{RhCl_3(PMe_2Ph)_3\}$ and excess potassium xanthate in refluxing ethanol was reported to give four products: $K\{RhCl_2(S_2CO)(PMe_2Ph)_2\}$ (I) (7%), $\{RhCl(S_2CO)(PMe_2Ph)_3\}$ (16%), $\underline{trans}$-$\{Rh(S_2CO)(S_2COEt)(PMe_2Ph)_2\}$ (17.5%), and its $\underline{cis}$-isomer (35%). The formulation of I was based on elemental analyses (found C, 34.9; H, 3.8%. I requires C, 35.1; H, 3.8%), qualitative detection of

chlorine by X-ray fluorescence, and conductivity measurements indicating a 1:1 electrolyte in nitromethane solution. A strong band in the infrared spectrum at 320 cm^{-1} was assigned to ν Rh-Cl, and peaks at 1640 and 1610 cm^{-1} to ν C-O (S_2CO^{2-}) [2]. Furthermore, ^1_1H n.m.r. studies in deuteroacetone showed a single triplet at τ=8.22, suggesting trans phosphine groups. The X-ray structural analysis reported here indicates that this compound is in fact trans-$K\{Rh(S_2CO)_2(PMe_2Ph)_2\}$ (II), which requires C, 35.9; H, 3.7%.

Experimental

Crystal data. $C_{18}H_{22}O_2KP_2RhS_4.3H_2O$, monoclinic, a = 18.798(4), b = 6.516(2), c = 23.924(5)Å, β = 113.91(1)°, V = 2697Å^3, D_m = 1.611, D_c = 1.617g/cm^3, M = 657, Z = 4, space group $P2_1/c$, F(000) = 1336, μ(Mo-Kα) = 12.2cm^{-1}.

Intensity data, Structure and Refinement. Intensity data were collected with Zr-filtered Mo-Kα radiation on a computer controlled NONIUS three circle diffractometer (θ-2θ scan). Of the 2480 attainable symmetry independent reflections within the limit θ<20°, 1648 reflections were 'observed' with intensities greater than three standard deviations, based on counting statistics. No absorption correction was applied. The structure was solved by a combination of Patterson and Direct methods[3]. From the Patterson synthesis, two independent Rh atoms were placed on special positions; these atoms only contribute to reflections in parity groups ggg and uuu. A Wilson-type plot for these reflections, and a Wilson-plot for the remaining reflections gave overall temperature parameters B_H and B_L for the rhodium atoms and for the remaining atoms, respectively. After substraction of the scaled heavy atom contribution from the observed value of the structure factors, normalized structure factors for the remaining light-atom structure were calculated. With the Sign Correlation Procedure[4], using a starting set of 50 reflections signed by the heavy atoms, 2 reflections to completely specify the origin, and 20 reflections with symbolic signs, 738 reflections were signed, and all symbols eliminated. A subsequent Fourier synthesis showed all light atoms except the hydrogens. The refinement was carried out by full-matrix least-squares, including anisotropic thermal parameters for the Rh, S, P and K atoms, and isotropic temperature factors for all other atoms, no attempt being made to locate hydrogen atoms. The least-

squares weights were w = sinθ/0.3 for sinθ<0.3, and w = 1 otherwise. The final refinement resulted in a residual R = 0.053 for all 'observed' reflections. Standard crystallographic calculations were carried out using XRAY70[5] as implemented at the Computing Centre of the University of Nijmegen.

Positional and thermal parameters are given in Tables 1 and 2, and the structure factor table is in Supplementary Publication No. SUP 0000.

Table 1. Fractional coordinates (x 10^4) with standard deviation for II.

Atom	Complex 1			Complex 2		
	x	y	z	x	y	z
Rh	0	0	0	5000	0	5000
S(1)	819(2)	2545(5)	648(2)	5032(2)	-1555(6)	5904(2)
S(2)	1304(2)	-1084(6)	249(2)	4879(2)	2712(6)	5623(2)
P	-65(2)	-1504(6)	872(2)	3660(2)	-725(6)	4578(2)
O	2300(6)	1696(17)	962(5)	4930(5)	1215(15)	6671(4)
C(1)	1603(8)	1132(22)	672(6)	4946(7)	888(20)	6162(6)
C(2)	-736(8)	-3670(24)	732(6)	3451(9)	-3429(26)	4651(7)
C(3)	856(8)	-2511(23)	1443(6)	3070(9)	597(24)	4914(7)
C(4)	-370(7)	229(21)	1325(5)	3165(7)	-188(21)	3757(6)
C(5)	181(8)	1561(21)	1765(6)	2809(9)	1713(24)	3578(7)
C(6)	-71(8)	2948(23)	2103(6)	2461(10)	2236(28)	2931(7)
C(7)	-855(9)	3034(25)	1996(7)	2504(11)	722(26)	2518(8)
C(8)	-1382(9)	1762(24)	1591(7)	2804(9)	-1110(27)	2697(7)
C(9)	-1158(8)	347(22)	1235(6)	3166(8)	-1616(24)	3336(6)
K	4287(2)	6980(5)	2123(2)			
O(3)	3826(6)	736(18)	1681(5)			
O(4)	4193(7)	3593(19)	2713(5)			
O(5)	2786(7)	5911(19)	1340(5)			

Discussion

Bond lengths and bond angles for the two independent complex ions are given in Tables 3 and 4, the numbering scheme adopted being given in Figure 1, and a projection of the unit cell along b in Figure 2.

Table 3. Bond lengths in the complex ions.

Bond	Complex 1	Complex 2
Rh-S(1)	2.364(4)Å	2.368(4)Å
Rh-S(2)	2.383(4)	2.382(4)
Rh-P	2.352(4)	2.351(4)
S(1)-C(1)	1.718(15)	1.739(15)
S(2)-C(1)	1.722(15)	1.720(15)
C(1)-O	1.264(16)	1.247(18)
P-C(2)	1.831(16)	1.827(19)
P-C(3)	1.838(12)	1.829(18)
P-C(4)	1.813(15)	1.834(12)
mean C-C (phenyl)	1.405(20)	1.399(20)

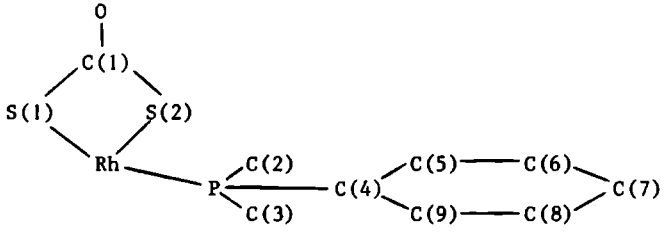


Figure 1. Numbering of the atoms in the two complex ions.

The rhodium atoms are both on crystallographic centres of symmetry with distorted octahedral coordination. In both complexes, the rhodium atom and the two dithiocarbonate ligands are essentially coplanar, the maximum deviation from the plane defined by the relevant atoms being 0.019Å (complex 1) and 0.015Å (complex 2). The only distortion of the dithiocarbonate ions from their ideal $\underline{C}_{2v}$ (mm) symmetry is the marginally significant difference between the angles S(1)-C(1)-O and S(2)-C(2)-O, also reflected in the Rh-S(1) and Rh-S(2) bond lengths, in both complexes. This is probably the result of the involvement of the oxygen atoms in hydrogen bonding and potassium coordination.

The dimethylphenylphosphine ligands have $\underline{C}_s$ (m) symmetry within the accuracy of the structure determination; the angle between the plane

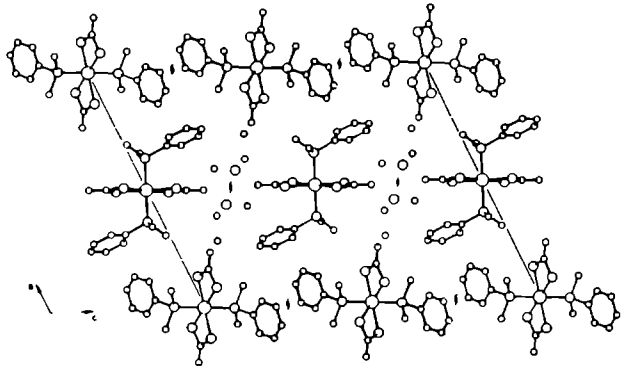


Figure 2. Projection of II along b. The large unconnected circles are potassium ions, while the small ones are water molecules.

defined by the phenyl ring and that defined by P, C(4), and the midpoint between C(2) and C(3) is 85° in complex 1 and 88° in complex 2. The essential difference between the two complex ions lies in the orientation of the phosphine group with respect to the rest of the complex. It is indicated by the pseudo-torsion angle C(4)-P-Rh-C(1), which is 90° in complex 1 and 150° in complex 2.

The potassium ions have a distorted trigonal prismatic coordination, given in Table 5. Each potassium ion shares an O(2) and an O(4) with one ion below and one above it about a screw axis. The hydrogen bonding pattern as well is associated with these columns about the screw axes at $\frac{1}{2}$, y, $\frac{1}{4}$ and $\frac{1}{2}$, y, $\frac{3}{4}$. Although the hydrogen atoms were not found, it is possible to assign unambiguously the sense of each hydrogen bond. Among the atoms found, there are no other 'intermolecular' contacts of less than 3.3Å.

The structure found for II agrees with the evidence used to assign it to I, except for the qualitative finding of chlorine, which appears to have been an impurity, and the assignment of an infrared band to a Rh-Cl stretch. As this band appears in a complex region of the spectrum, it could be assigned otherwise, probably to a Rh-S stretch.

Table 4. Bond angles in the complex ions.

Angle	Complex 1	Complex 2
S(1)-Rh-S(2)	73.35(13)°	73.68(13)°
S(1)-Rh-P	87.96(13)	86.15(13)
S(2)-Rh-P	94.15(13)	93.47(13)
Rh-S(1)-C(1)	87.9(3)	87.8(3)
Rh-S(2)-C(1)	87.3(3)	87.7(3)
S(1)-C(1)-S(2)	111.0(7)	110.8(8)
S(1)-C(1)-O	123.2(11)	123.0(10)
S(2)-C(1)-O	125.8(11)	126.1(11)
Rh-P-C(2)	116.1(3)	112.8(3)
Rh-P-C(3)	115.9(3)	116.9(3)
Rh-P-C(4)	114.1(3)	114.3(3)
C(2)-P-C(3)	103.1(5)	102.7(5)
C(2)-P-C(4)	103.5(5)	104.5(5)
C(3)-P-C(4)	102.2(5)	104.2(5)
P-C(4)-C(5)	120.7(11)	118.1(12)
P-C(4)-C(9)	120.6(11)	120.4(12)
mean C-C-C (phenyl)	120.0(13)	120.0(15)

We thank Drs. Beurskens, Cole-Hamilton, Noordik, and Stephenson for the sample and for much helpful discussion. One of us (Th.E.M.v.d.H.) acknowledges the support of the Netherlands Foundation for Pure Research.

References

[1]D.J. Cole-Hamilton and T.A. Stephenson, J.C.S. Dalton, 1974, 1818.

[2]J.M. Burke and J.P. Fackler Jr., Inorg.Chem., 1972, 11, 2744.

[3]P.T. Beurskens and J.H. Noordik, Acta Cryst., 1971, 27, 187.

[4]P.T. Beurskens, Acta Cryst., 1964, 17, 462, and 'Lecture Notes', N.A.T.O. Advanced Study Institute on Direct Methods in Crystallography, Erice, 1974.

[5]J.M. Stewart et al., 'The XRAY system', Technical Report of the University of Maryland computer Science Center, 1970.

CHAPTER 14

PROCEDURE DIRDIF.B

A description of the procedure DIRDIF.B is given in the following reprint:

Acta Cryst (1975) A31, 813

The Application of Direct Methods to Centrosymmetric Structures containing Heavy Atoms. II

BY R O GOULD

Edinburgh University, Department of Chemistry, West Mains Road, Edinburgh, Scotland

AND TH. E M VAN DEN HARK AND PAUL T BEURSKENS*

Crystallography Laboratory, Toernooiveld, Nijmegen, The Netherlands

(*Received* 12 *May* 1975, *accepted* 23 *June* 1975)

Direct methods are applied to the difference structure factors for a structure containing one or more heavy atoms in known positions For those reflexions whose sign is determined by the heavy atoms, the known heavy-atom contribution is subtracted from the observed structure factor to obtain the magnitude and the sign of the light-atom contribution The signs of the reflexions that do not have an appreciable heavy-atom contribution are found using the $\sum_1$ sign relationship For reflexions with an intermediate heavy-atom contribution the sign and magnitude ambiguity is also solved by the $\sum_2$ formula Thus it is possible to maximize the number of correct signs, and correct some of the amplitudes, before a difference Fourier map is calculated

Introduction

In paper I (Beurskens & Noodik, 1971) direct methods were used to solve the phase problem for those special cases where the heavy atoms are on special positions, such that the heavy atoms do not contribute to several reflexion parity groups In those cases the crystal structure is not determined by the positions of the heavy atoms only one or two phases have to be chosen in order to specify the origin fully †

The present paper deals with the general case the positions of the heavy atoms completely determine the structure, the phase problem is solved in principle the positions of the remaining 'light' atoms can be found by standard Patterson and Fourier techniques The time and effort necessary for finding the light-atom structure is reduced by the present procedure, especially when the heavy atoms are marginally sufficient to solve the phase problem

Define

$|F_{obs}|$ is the observed structure factor
F_H is the calculated contribution of the known part of the structure ('heavy' atoms).
F_L is the contribution of the remaining part of the structure ('light' atoms)
$S(F)$ is the sign of F.

* Author for correspondence

† A Fortran program for the execution of this procedure, *DIRDIF A*, is now available on request

In the normal procedure, difference coefficients,

$$\Delta F = S(F_H)\ \ (|F_{obs}| - |F_H|)$$

are calculated and accepted as an estimate for F_L Only in very favorable circumstances can the complete light-atom structure be unambiguously deduced from the Fourier synthesis based on these coefficients In the present procedure, a $\sum_2$ refinement procedure is used to convert input ΔF values to correct F_L values

Two categories of reflexions may be distinguished (see Fig 1)

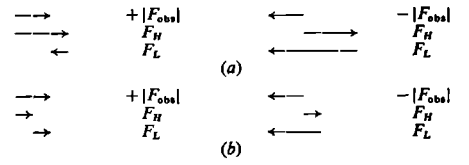


Fig 1 Two categories of reflexions Drawings for a given $|F_{obs}|$ value, and $S(F_H) = +$ See text

(*a*) $|F_H| > |F_{obs}|$ In this case [Fig 1(*a*)] the sign of F_L is known to be opposite to the sign of F_H Nevertheless, the absolute value of F_L is not known for certain as the sign of F_{obs} is not known The more probable value of $|F_L|$ is the smaller of the two possibilities If this is correct, then $F_L = \Delta F$ There will in general be very few reflexions in this category for which the larger magnitude is correct *and* significantly different from the smaller, no attempt is made to find these reflexions

(*b*) $|F_{obs}| > |F_H|$ In this case [Fig 1(*b*)] neither sign nor absolute value of F_L is known, but sign and absolute value are correlated Again the smaller of the two possible $|F_L|$ values is the more probable one, corresponding to $F_L = \Delta F$ If ΔF is sufficiently large (see below) then this reflexion will take part in our $\sum_2$ refinement procedure, and when it is found that the sign of F_L is wrong, we not only change the sign but also the absolute value of F_L In the special case where the heavy-atom contribution is very weak, if not zero, ΔF cannot be calculated, nevertheless the absolute value of F_L is known, and the sign of F_L may be calculated by the $\sum_2$ formula

Description of the procedure

The procedure consists of the following steps calculation of a 'two-dimensional' Wilson plot, the calculation of 'dual' E values for the light-atom structure, the determination and refinement of signs by the $\sum_2$ formula, and the selection of Fourier coefficients.

Symbols used

$I = K\ |F_{obs}|^2$, observed intensity on a relative scale K

F_H^0 calculated structure factor on absolute scale for the known heavy atoms with a temperature factor of $B_H = 0\ 0$

f_H, f_L, f scattering factors for a heavy, light, or any atom

B_H, B_L, B_{ov} overall temperature factors for the heavy atoms, the unknown part of the structure, or the whole structure

E_H, E_L normalized structure factors for a structure consisting of heavy or light atoms only

$\sum^H, \sum^L, \sum$ summation over the heavy, light, or all atoms of a unit cell

$S = \sin\theta/\lambda$

Wilson plot

We have stressed (Paper I) the importance of a well designed Wilson-plot routine to obtain the best possible estimates for the $|F_L|$ values and to avoid disasters during the automatic execution of the procedure The 'local' intensity average for a partially known structure is [equation (2) in Paper I, Parthasarathy (1966)]

$$\langle I\rangle_{\mathbf{h}} = K[\langle \textstyle\sum^L f_L^2 \exp(-2B_L S^2)\rangle_{\mathbf{h}} + \langle |F_H^0|^2 \exp(-2B_H S^2)\rangle_{\mathbf{h}}] \quad (1)$$

The average is taken over reflections **h** within a given interval

Initially, setting $B_L = B_H = B_{ov}$, and applying (1) in ranges of S, a one-dimensional Wilson plot is calculated In Paper I the value of B_L was found by a Wilson plot on those reflexions that had no heavy-atom contributions, thereafter B_H was found by a 'difference' Wilson plot

It is often possible to refine simultaneously the temperature factors of the heavy and light atoms separately The main restriction is that the heavy atom must not lie in a position (such as a centre of symmetry in space group $P\bar{1}$), where it contributes uniformly to all reflexions Less accurate results may be expected if there are a number of moderately heavy atoms (*e g* S or P) grouped around the known atoms, but this should normally give little trouble The procedure, adopted in our *DIRDIF B* program, is a non-linear iterative one, and depends on the reflexions being distributed over a two-dimensional array in ranges of S and of $|E_H|$, where

$$|E_H|^2 = |F_H^0|^2 / \textstyle\sum^H f_H^2$$

Define

$$G_{obs} = \langle I / \textstyle\sum f^2\rangle_{\mathbf{h}}$$
$$G_H = K\langle |F_H^0|^2 \exp(-2B_H S^2) / \textstyle\sum f^2\rangle_{\mathbf{h}}$$
$$G_L = K\langle \textstyle\sum^L f_L^2 \exp(-2B_L S^2) / \sum f^2\rangle_{\mathbf{h}},$$

where the averaging is done over the appropriate range of S and $|E_H|$ The quantity to be minimized is then

$$\sum_{\text{ranges}} (G_{obs} - G_H - G_L)^2$$

and the parameters to be refined are K, B_H and B_L Refinement normally takes three cycles

Normalized structure factors

For every reflexion two possible values for E_L, the normalized structure factor of the light-atom structure are calculated:

$$E_1=[S(F_H)\,.\,|F_{obs}|-F_H]/(\varepsilon\sum {}^{L}f_L^2)^{1/2}\exp(-B_LS^2)$$
$$E_2=[-S(F_H)\,.\,|F_{ob.}|-F_H]/(\varepsilon\sum {}^{L}f_L^2)^{1/2}\exp(-B_LS^2) \quad (2)$$

where $|F_{obs}|=(I/K)^{1/2}$, $F_H=F_H^0\exp(-B_HS^2)$, and ε is the well-known correction for symmetry enhancement (Hauptman & Karle, 1953). Always $|E_1|\le|E_2|$; for most reflexions $|E_1|<|E_2|$, and E_1 is the most probable value. It is the aim of our procedure to find reflexions where the less probable value, E_2, is the correct value.

Let us define P_1 as the probability that the *sign* of E_1 is correct. For reflexions in category *b*, the probability that the *magnitude* of E_1 is correct is also given by P_1. This probability can be calculated from the magnitudes of E_1 and E_2, using the distribution of E values in centrosymmetric space groups: $P(E)=(2\pi)^{-1/2}\exp[-\frac{1}{2}E^2]$ (Hauptman & Karle, 1953). This gives (Woolfson, 1956):

$$P_1=\exp(-\tfrac{1}{2}E_1^2)/[\exp(-\tfrac{1}{2}E_1^2)+\exp(-\tfrac{1}{2}E_2^2)] \quad (3)$$

(see Table 1) and the weight to be assigned to the value E_1 is taken as:

$$W_1=(2P_1-1)^2 \quad (4)$$

(Note: $|E_1|\le|E_2|$. $0{\cdot}5\le P_1\le 1$, thus: $0\le W_1\le 1$). The special case $F_H\simeq 0$ leads to $E_1\simeq -E_2$ with $P_1\simeq P_2\simeq 0{\cdot}5$, which does not need special treatment.

For reflexions in category *a*, E_1 and E_2 have the same sign, and this sign is known and will be used to determine other signs. As explained above, we do not try to find the correct absolute value, and for safety reasons we have to accept the lower $|E|$ values. Therefore E_1 is accepted, with P_1 equal to 1·0:

$$\text{for}\qquad |F_H|>|F_{obs}|:\quad P_1=1{\cdot}0\,,\quad W_1=1{\cdot}0\,. \quad (5)$$

The special case $F_{obs}\simeq 0$, however, may easily lead to severe practical problems, especially for high-order reflexions. These are the so-called 'less-than' or 'unobserved' reflexions. The $|F_{obs}|$ value is very badly determined and, to avoid overestimating the E_L values, the E_1 and E_2 values are decreased by one or two standard deviations. It should be noted that the unobserved reflexions may play an important role in this procedure.

Σ_2 formula

When a sufficient number of signs is known, the $\sum_2$ formula (Hauptman & Karle, 1953) will easily lead to refinement of probable signs, and to a calculation of new signs. The weighted $\sum_2$ formula may be given as:

$$S(E_{\mathbf{h}})\simeq S(\textstyle\sum_{\mathbf{k}} W_{\mathbf{k}}W_{\mathbf{h}-\mathbf{k}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}) \quad (6)$$

where $E_{\mathbf{k}}$ is E_L for the reflexion $\mathbf{k}$, and $W_{\mathbf{k}}$ is the weight for $E_{\mathbf{k}}$. The summation is over all available terms.

For weights, defined according to (4), it is justified to generalize the formula for the probability that (6) gives the correct result (Cochran & Woolfson, 1955):

$$P_S=\tfrac{1}{2}+\tfrac{1}{2}\tanh\left[\frac{\sigma_3}{\sigma_2^{3/2}}\textstyle\sum_{\mathbf{k}}W_{\mathbf{k}}W_{\mathbf{h}-\mathbf{k}}E_{\mathbf{k}}E_{\mathbf{h}}E_{\mathbf{h}-\mathbf{k}}\right] \quad (7)$$

with $\sigma_n=\sum^{L}Z^n$, Z being the atomic number of the light atoms. Analogous to (4) we define for the weight of the $\sum_2$ result:

$$W_S=(2P_S-1)^2\,. \quad (8)$$

These formulae are applied as follows:

(i) Reflexions with $|E_1|$ less than some minimum value (say: $E_{min}=0{\cdot}9$) are not used.

(ii) The reflexions of category *a* and some of the reflexions of category *b* (*i.e.* those with P_1 nearly equal to unity) are not recalculated; their E_1 values serve as input to the $\sum_2$ formula.

(iii) The remaining reflexions serve as input as well as output to equation (6).

Any resulting sign, $S(E_{\mathbf{h}})$ in (6), is compared with the input sign $S(E_1)$ for the reflexion $\mathbf{h}$. In case these signs are not equal, and if the new sign is more reliable (*i.e.* if $W_S>W_1$), then E_1 is replaced by E_2 for this reflexion, and W_S is taken as its weight.

Convergence is achieved by repeating these calculations three or four times. Most of the reflexions still have $E_L=E_1$, but some are changed to $E_L=E_2$. Relatively more reflexions have $E_L=E_2$ for structures where the known 'heavy' atoms do not fix the signs of the dif-

Table 1. *Probability P_1 for dual E_L values*

$\lvert E_2\rvert=$	0·2	0·4	0·6	0·8	1·0	1·2	1·5	2·0	2·5	3·0	3·5	4·0	5·0
$\lvert E_1\rvert=0{\cdot}2$	0·50	0·52	0·54	0·57	0·62	0·67	0·75	0·88	0·957	0·989	0·998	1·000	1·000
$\lvert E_1\rvert=0{\cdot}4$		0·50	0·52	0·56	0·60	0·66	0·74	0·87	0·954	0·988	0·998	1·000	1·000
$\lvert E_1\rvert=0{\cdot}6$			0·50	0·54	0·58	0·63	0·72	0·86	0·950	0·987	0·997	1·000	1·000
$\lvert E_1\rvert=0{\cdot}8$				0·50	0·55	0·60	0·69	0·84	0·943	0·985	0·997	1·000	1·000
$\lvert E_1\rvert=1{\cdot}0$					0·50	0·56	0·65	0·82	0·932	0·982	0·996	0·999	1·000
$\lvert E_1\rvert=1{\cdot}2$						0·50	0·60	0·78	0·917	0·978	0·996	0·999	1·000
$\lvert E_1\rvert=1{\cdot}5$							0·50	0·71	0·88	0·967	0·993	0·999	1·000
$\lvert E_1\rvert=2{\cdot}0$								0·50	0·76	0·924	0·984	0·998	1·000
$\lvert E_1\rvert=2{\cdot}5$									0·50	0·80	0·952	0·992	1·000
$\lvert E_1\rvert=3{\cdot}0$										0·50	0·84	0·971	1·000
$\lvert E_1\rvert=3{\cdot}5$											0·50	0·87	0·998
$\lvert E_1\rvert=4{\cdot}0$												0·50	0·989
$\lvert E_1\rvert=5{\cdot}0$													0·50

ference coefficients well, either because they are in pseudo-special positions, or because they are not sufficiently heavy relative to the other atoms.

Fourier coefficients

Practically all reflexions used in the Σ_2 calculations, have acceptable E_L values, some reflexions with probabilities (P_1 and P_5) less than some minimum value (say. $P_1 < 0{\cdot}8$ and $P_5 < 0{\cdot}97$) are rejected.

In addition to these results we carefully select the weaker E_L values, that is the reflexions with $|E_L| < E_{min}$. In category *a* all such reflexions have known signs and can be used as Fourier coefficients. In category *b* many such reflexions have E_L values with acceptable probabilities (say. $P_1 > 0{\cdot}8$) and can therefore be used.

For all accepted reflexions the E_L value is multiplied by a weighting factor (W_1 or W_5) and – if wanted – the E_L values are converted to F_L values. A Fourier synthesis based on these E_L or F_L values should give the complete light-atom structure.

Numerical results*

The present procedure has been tested on four known structures (1–4, Table 2) and has actually been used in the determination of three crystal structures (5–7, Table 2). In all these cases only one metal atom has been accepted as a 'heavy atom'. Of course, the final results will be better if one can add more atoms to the 'known part' of the structure, say one K or one or two *S* atoms in some of the structures listed in Table 2.

Three of the structures (2, 3 and 5) have been analysed in more detail; the results are listed in Tables 3 through 5. In general, some 10 to 20% of the strong reflexions ($|E_L| > 1{\cdot}0$) will change their signs (*i.e.* $E_L = E_2$), including some reflexions with very different $|E_1|$ and $|E_2|$ values. Moreover, a comparable number of strong reflexions with a weak heavy-atom contribution ($|E_L| > 1{\cdot}0$, $|E_H| < 0{\cdot}5$), often excluded from a difference Fourier synthesis, will now have a calculated sign. In our test cases all of the non-hydrogen atoms were found except, in some cases, some of the strongly vibrating carbon atoms. (For example, for structure 5 a difference Fourier map, signed by the Rb atom only, showed 31 of the 42 atoms of the molecule among the top 42 peaks, while a difference Fourier calculated from the *DIRDIF B* output gave the complete structure from the top 42 peaks.)

* A Fortran program, *DIRDIF B*, is available on request.

Table 3. *Classification of reflexions according to magnitude $|E_1|$ and probability P_1*

Structure no.	2	3	5
Number of reflexions with $\|E_1\| < 0{\cdot}1$*	454	197	510
Reflexions with $0{\cdot}1 \le \|E_1\| < E_{min}$			
$P_1 < 0{\cdot}7$*	474	275	1032
$P_1 \ge 0{\cdot}7$†	2444	624	1998
Reflexions with $\|E_1\| \ge E_{min}$			
category *a*	185	124	86
category *b*, $P_1 \ge 0{\cdot}9999$	144	88	140
category *b*, $P_1 < 0{\cdot}9999$	538	350	423
Total number of reflexions (E_{min})	4239 (1·1)	1658 (0·8)	4189 (1·1)

* Reflexions in these classes will not be used for the calculation of a difference Fourier map.

† These reflexions are analysed in Table 5.

Comment about structure no. 4: the Mo atom is situated on a twofold axis, with a *y* coordinate near to $\frac{1}{8}$ ($y = 0{\cdot}1264$; deviation from $y = \frac{1}{8}$: 0·025 Å). Therefore a quarter of all reflexions have almost no Mo contribution to the structure factor. Nevertheless, the origin is completely fixed by the position of the Mo atoms, and our procedure works without any trouble at all.

Comment about structure no. 5: the *y* coordinate of the Rb atom is near to $\frac{1}{4}$ ($y = 0{\cdot}262$, deviation from $y = \frac{1}{4}$. 0·12 Å). Therefore, one half of all reflexions have almost no Rb contribution to the structure factor. Taking $y = 0{\cdot}25$, the origin is not fixed, and procedure *DIRDIF A* (Paper I) could have been used, with, of course, an inaccurate Wilson plot for the weak reflexions. The origin is fixed by taking $y = 0{\cdot}26$ or 0·27 and the present procedure gives the structure completely.

Table 2. *Test structures for DIRDIF. B*

N' = number of symmetry-independent atoms per unit cell (hydrogens excluded), HA = known part of the structure; perc = percentage of scattering power represented by HA, N_{ref} = number of symmetry-independent reflexions.

No	Structure	Space group	*Z*	*N'*	HA	Perc	N_{ref}	References
1*	$AuS_4N_3C_{13}H_{18}$	*Pbca*	8	21	Au	30%	3283	Noordik, Hummelink & van der Linden (1973)
2	$CoK_3O_6N_6C_6H_9$ $6\frac{1}{2}H_2O$	$P\bar{1}$	2	31	Co	9%	4239	Birker, Smits, Bour & Beurskens (1973)
3	$NiS_4N_4C_{14}H_{18}$	$P2_1/c$	4	21	Ni	13%	1658	Hendriks, Bosman & Beurskens (1974)
4	$MoS_8N_4C_{16}H_{24}$	$C2/c$	4	17	Mo	12%	1297	van der Aalsvoort & Beurskens (1974)
5	$RbO_{10}C_{32}H_{54}$	$C2/c$	8	43	Rb	10%	4189	Mooy, Degens, Noordik & van den Hark (1975)
6	$RhClS_3PN_2C_{24}H_{27}$ $CHCl_3$	$P2_1/c$	4	36	Rh	12%	2905	Bosman & Gal (1975)
7	$MoS_6N_5C_{13}H_{10}$ $N(C_2H_5)_4$	$P2_1/c$	4	42	Mo	10%	3570	Bosman & Nieuwpoort (1975)

* The preliminary procedure, applied to this structure (Noordik, 1971) has now been improved considerably.

Table 4 *Final results for reflexions signed by the $\sum_2$ formula (number of reflexions with $E_L = E_1$ and $E_L = E_2$ for different ranges of $|E_2| - |E_1|$, for all reflexions in category b with $P_1 < 0.9999$)*

The number of incorrectly signed reflexions is given in parentheses

Range of $(\|E_2\| - \|E_1\|)$	Structure 2		Structure 3		Structure 5	
	$E_L = E_1$	$E_L = E_2$	$E_L = E_1$	$E_L = E_2$	$E_L = E_1$	$E_L = E_2$
0.0–0.5*	66	37 (1)	70 (1)	53 (3)	102 (1)	56
0.5–1.0	63	20	38	18 (2)	60	10
1.0–1.5	61	7	49	10 (1)	69	4
1.5–2.0	97	2	39	5	41	0
2.0†	184	1	68	0	79	2
Total	471	67 (1)	264 (1)	86 (6)	351 (1)	72
Determined with probability ≥ 0.97	532		334 (2)		396 (1)	
Determined with probability < 0.97	6 (1)		16 (5)		27	

* Most of these reflexions have probabilities P_1 in the range 0.50–0.75
† All of these reflexions have $P_1 > 0.98$

Table 5 *Analysis of weak reflexions, for which the signs are not refined by the $\sum_2$ formula (all reflexions with $0.1 \leq |E_1| < E_{min}$, $P_1 \geq 0.7$)*

Structure no	2	3	5
Probability range (P_1)			
0.7—0.8	151 (46)	60 (18)	167 (58)
0.8—0.9	143 (22)	47 (9)	165 (41)
0.9—0.95	121 (14)	31 (4)	89 (11)
0.95–0.999	450 (18)	86 (7)	281 (29)
> 0.999*	1579 (58)	400 (24)	1296 (66)

* This range includes reflexions of category *a* (with $P_1 = 1.0$) The majority of reflexions with wrong signs in this range had very small $|E_1|$ values ($|E_1| < 0.2$), a wrong sign may then be caused by small errors in scale and temperature factors or by errors in the measured intensity
The number of incorrect signs is given between parentheses

Conclusions

The procedure described above proved to be useful in routine structure analysis of centrosymmetric heavy-atom compounds The calculations are performed by an automatic computer program for the use of which no direct methods experience is required We expect to be able to apply an analogous procedure to non-centrosymmetric structures as well

One of us (Th E M van den Hark) acknowledges support of the Dutch Foundation for Pure Research, ZWO/FOMRE

References

AALSVOORT, J G M VAN DER & BEURSKENS, P T (1974) *Cryst Struct Commun* **3**, 653–656
BEURSKENS, P T & NOORDIK, J H (1971) *Acta Cryst* A**27**, 187–188
BIRKER, P J M W L, SMITS, J M M, BOUR, J J & BEURSKENS, P T (1973) *Recueil*, **92**, 1240–1248
BOSMAN, W P & GAL, A W (1975) To be submitted to *Cryst Struct Commun*
BOSMAN, W P & NIEUWPOORT, A (1975) To be submitted to *Cryst Struct Commun*
COCHRAN, W & WOOLFSON, M M (1955) *Acta Cryst* **8**, 1–12
HAUPTMAN, H & KARLE, J (1953) *Solution of the Phase Problem I The Centrosymmetric Crystal* A C A Monograph No 3 Pittsburgh Polycrystal Book Service
HENDRIKS, H M, BOSMAN, W P & BEURSKENS, P T (1974) *Cryst Struct Commun* **3**, 447–450
MOOY, J J, DEGENS, H M L, NOORDIK, J H & VAN DEN HARK, Th E M (1975) To be published
NOORDIK, J H (1971) Thesis, Univ of Nijmegen
NOORDIK, J H, HUMMELINK, Th W & VAN DER LINDEN, J G M (1973) *J Coord Chem* **2**, 185–191
PARTHASARATHY, S (1966) *Z Kristallogr* **123**, 27–50
WOOLFSON, M M (1956) *Acta Cryst* **9**, 804–810

Reprinted from *Acta Crystallographica*, Vol A 31, part 6, November 1975

PRINTED IN DENMARK

CHAPTER 15

CRYSTAL STRUCTURE OF 'RUBIF'

In this chapter part of the paper on the 'RUBIF' structure is reproduced as an example of the application of the DIRDIF.B procedure. For brevity we have omitted: thermal parameters, n.m.r. and magnetic properties and the last part of the discussion of the crystal structure. Some numerical data concerning the application of the DIRDIF.B procedure are given in the preceding chapter.

Reprinted with permission from the Journal of the American Chemical Society (1976) 98, 680.

Molecular and Magnetic Structure of the Paramagnetic Ion Pair Bis(Tetraglyme)Rubidium$^+$ Biphenyl$^-$, $Rb^+\{CH_3O(CH_2CH_2O)_4CH_3\}_2C_{12}H_{10}^-$

J.J. Mooij, A.A.K. Klaassen, E. de Boer[x],
H.M.L. Degens, Th.E.M. van den Hark and J.H. Noordik

Contribution from the Department of Physical Chemistry[x] and the Crystallography Laboratory, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands

Abstract

Single crystals of the paramagnetic ion pair bis(tetraglyme)rubidium$^+$biphenyl$^-$, $Rb^+\{CH_3O(CH_2CH_2O)_4CH_3\}_2C_{12}H_{10}^-$ have been synthesized. The crystal structure was determined at 100°K from three dimensional X-ray data collected by counter methods. The crystals belong to the space group C2/c with $\underline{a}$ = 30.68 (3), $\underline{b}$ = 9.79 (1), $\underline{c}$ = 23.71 (2) Å, β = 103.34 (6)°, Z = 8 and D_x = 1.31 g/cm^3. Each rubidium ion is spherically surrounded by ten oxygen atoms of the solvent molecules, leading to a solvent separated ion pair structure. The phenyl rings of the biphenyl anion have a dihedral angle of 9.4°. The magnetic properties have been investigated by means of esr, nmr and the foner vibrating sample magnetometer. It is shown that the results of both the paramagnetic susceptibility and the angular dependent linewidth variation of the ex-

change narrowed esr line are consistent with the observed dimeric structure of the biphenyl anions. The susceptibility data and the temperature dependent linewidth measurements indicate an exchange coupling in the dimer with a singlet groundstate separated from the higher triplet state by $2J = 16.7\ cm^{-1}$.

Introduction

During the last decades much work has been carried out on the negative ions of aromatic hydrocarbons, prepared by reduction of the neutral molecules with alkali metals[1]. Especially esr and nmr experiments have provided much information on the structure of the alkali radical ion pairs in solution. Based on these experiments two different types of ion pairs have been distinguished, namely "tight" or "contact" ion pairs and "loose" or "solvent separated" ion pairs[1]. To shed more light on the structure of the ion pairs it would be of interest to carry out experiments on single crystals of the ion pairs. In 1970 Canters et al.[2] reported the first successful preparation of single crystals of the paramagnetic ion pairs of alkali biphenyl. A few years later Brooks et al.[3] published the crystal structures of some diamagnetic ion pairs, namely, triphenylmethyllithium tetramethylethylenediamine, fluorenyllithium bisquinuclidine and the bis{(tetramethylethylenediamine)lithium} naphthalene dianion. In all these complexes the lithium ion is directly coordinated to the aromatic moiety and from the outside coordinated to the nitrogen atoms of the solvent molecule, so that these complexes belong to the category of "contact" ion pairs. Recently the crystal structure of the potassium salt of 1,3,5,7-tetramethylcyclooctatetraene dianion has been published by Goldberg at al.[4] Noordik et al.[5] have published the crystal structures of the potassium and rubidium salts of the dianion of unsubstituted cyclooctatetraene (COT). All these crystals contained molecules of the solvent, diglyme ($CH_3O(CH_2CH_2O)_2CH_3$), from which the crystals were prepared. The crystal structures reveal that the alkali ions are located above the center of the planar eight membered ring; the distances to the center of the ring are equal to the sum of the radius of the cation and the half thickness of the COT dianion. In the tetramethyl substituted compound the diglyme molecules are coordinated to the potassium ions through all three oxygen atoms; in the unsubstituted COT

compounds two different types of alkali ions can be discerned, one type is coordinated only by COT rings, whereas the other is coordinated on one side to the oxygen atoms of diglyme and on the other side by a COT ring. Therefore the single crystals of the alkali COT complexes can also be classified as contact ion pairs. It is interesting to note that Cox et al.[6] arrived at the same conclusion from nmr experiments on dipotassium cyclooctatetraene dissolved in diglyme.

In this paper we report the first crystal structure of a paramagnetic ion pair, viz., the rubidium biphenyl (RbBp) ion pair. The crystals prepared from tetraglyme ($CH_3O(CH_2CH_2O)_4CH_3$) have the composition RbBp. 2 Ttg (Ttg = tetraglyme) and belong to the category of solvent seperated ion pairs, since the Rb cation is completely surrounded by two tetraglyme molecules and coordinated to the tetraglymes through the ten oxygen atoms.

The magnetic properties of the crystals have been investigated by esr, nmr and with the foner vibrating sample magnetometer. The esr spectrum consists of one single exchange narrowed line. Its linewidth depends on the orientation of the magnetic field with respect to the crystal axes. With the theory of Anderson and Weiss[7] and Van Vleck[8] a satisfactory explanation could be given for this linewidth variation. Susceptibility measurements revealed an antiferromagnetic coupling in the crystals. With the aid of a singlet triplet model a satisfactory account could be given for the temperature dependence of the susceptibility.

Experimental section

Preparation of the crystals

1.0 M solutions of the RbBp salt were prepared under high vacuum using standard techniques[2]. From these solutions single crystals were obtained by slowly cooling down the solutions to about 10°C, 1 degree per hour. The crystals were mounted in thin glass capillaries in a He atmosphere using a home-built glove box, since the crystals are sensitive to air and moisture. In view of the low melting point (45°C) of the crystals these manipulations were carried out at a temperature of about -20°C.

Structure determination

$Rb^{+}\{CH_3O(CH_2CH_2O)_4CH_3\}_2C_{12}H_{10}^{-}$, unit cell dimensions and intensities were obtained from a crystal of approximate dimensions 0.54 x 0.18 x 0.20 mm at 100°K, with a computer controlled NONIUS-CAD-3 diffractometer (Zr-filtered MoKα radiation, λ = 0.71069Å, θ-2θ scan). Systematic absences are hkl for h + k odd, h0l for l odd and 0k0 for k odd, these are consistent with the space groups C2/c and Cc. The structure was successfully refined in the centrosymmetric space group C2/c. Unit cell dimensions are: $\underline{a}$ = 30.68(3), $\underline{b}$ = 9.79(1), $\underline{c}$ = 23.71(2)Å, β = 103.34(6)°, V = 6909Å^3, D_x = 1.31g/cm^3, Z = 8. Linear absorption coefficient μ = 15.7 cm^{-1}.

Of the 6097 symmetry independent reflections up to θ = 25°, 3573 reflections were observed above background {$I>3\sigma_c(I)$, $\sigma_c(I)$ derived from counting statistics}. The observed intensities were corrected for Lorentz and polarization effects and absorption. The latter corrections were calculated according to the Busing and Levy[9] scheme; 4 x 4 x 4 volume fragments and 7 bounding planes were taken into account. The unobserved reflections were not used.

The structure was solved by a combination of Patterson and direct methods techniques. From a Patterson synthesis the position of the rubidium cation was determined. The remaining light atom structure was then solved by direct methods from the known heavy atom position using the automatic program DIRDIF.B[10]. The atomic coordinates and the anisotropic thermal parameters of all non-hydrogen atoms were refined by least-squares methods. The function that was minimized was $\Sigma\, w\{|F_o| - K|F_c|\}^2$, with $w^{-1} = \sigma(F_o)^2 + (0.05|F_o|)^2$. The hydrogen atoms were placed at calculated positions. Hydrogen atoms of the biphenyl anion were placed on the bisector of the C-CH-C angle (CH is the carbon atom to which the hydrogen atom is attached) at a C-H distance of 1.084Å. Hydrogen atoms attached to the end carbon atoms of the tetraglyme molecules were placed staggered with respect to the tetraglyme chain at a C-H distance of 1.101Å and H-C-H angles of 109.47°. The remaining hydrogen atoms of the tetraglyme molecules were placed in the plane bisecting the O-CH-C angle, at a distance of 1.073Å and with an H-C-H angle of 109.47°. All hydrogen atoms were included in the structure factor calculations with a temperature factor equal to the isotropic equivalent of their parent carbon atoms, but they were not refined. Final conven-

Table 1 Final Positional Parameters for the Nonhydrogen Atoms with Standard Deviation in Parentheses.

	x	y	z
Rb	0.12828(2)	0.26209(6)	0.11064(2)
Tetraglyme I			
C(13)	0.0752(3)	-0.0103(8)	-0.0125(3)
O(1)	0.0622(2)	0.1099(5)	0.0142(2)
C(14)	0.0327(3)	0.1921(7)	-0.0283(3)
C(15)	0.0210(2)	0.3171(8)	0.0025(3)
O(2)	0.0577(2)	0.4028(5)	0.0262(2)
C(16)	0.0744(3)	0.4764(8)	-0.0168(3)
C(17)	0.1102(2)	0.5730(7)	0.0136(3)
O(3)	0.1492(1)	0.4941(4)	0.0410(2)
C(18)	0.1861(3)	0.5799(7)	0.0642(3)
C(19)	0.2260(2)	0.4907(8)	0.0910(3)
O(4)	0.2178(1)	0.4348(5)	0.1431(2)
C(20)	0.2513(2)	0.3374(8)	0.1694(4)
C(21)	0.2427(2)	0.2917(7)	0.2258(3)
O(5)	0.2008(2)	0.2211(4)	0.2159(2)
C(22)	0.1897(3)	0.1887(8)	0.2694(3)
Tetraglyme II			
C(23)	0.1461(2)	0.5421(7)	0.2213(3)
O(6)	0.1126(2)	0.5291(5)	0.1672(2)
C(24)	0.0675(2)	0.5368(7)	0.1738(3)
C(25)	0.0538(2)	0.4167(7)	0.2052(3)
O(7)	0.0538(1)	0.2968(4)	0.1704(2)
C(26)	0.0383(2)	0.1815(7)	0.1970(3)
C(27)	0.0457(2)	0.0557(7)	0.1656(3)
O(8)	0.0931(1)	0.0376(4)	0.1726(2)
C(28)	0.1045(2)	-0.1019(7)	0.1634(3)
C(29)	0.1535(2)	-0.1074(6)	0.1631(3)
O(9)	0.1592(1)	-0.0428(5)	0.1112(2)
C(30)	0.2047(2)	-0.0373(6)	0.1077(3)
C(31)	0.2057(2)	0.0138(7)	0.0479(3)

Table 1 (continued)

O(10)	0.1904(1)	0.1525(4)	0.0432(2)
C(32)	0.1826(2)	0.1992(7)	-0.0154(3)
Biphenyl anion			
C(1)	0.1217(2)	0.4674(6)	0.3666(3)
C(2)	0.0934(2)	0.3487(7)	0.3611(3)
C(3)	0.1098(2)	0.2200(7)	0.3776(3)
C(4)	0.1559(2)	0.1987(7)	0.4021(3)
C(5)	0.1842(2)	0.3125(7)	0.4084(3)
C(6)	0.1682(2)	0.4426(7)	0.3910(3)
C(7)	0.1044(2)	0.6017(6)	0.3493(2)
C(8)	0.0573(2)	0.6284(6)	0.3331(3)
C(9)	0.0406(2)	0.7569(7)	0.3159(3)
C(10)	0.0690(2)	0.8678(7)	0.3132(3)
C(11)	0.1155(2)	0.8435(7)	0.3299(3)
C(12)	0.1325(2)	0.7151(6)	0.3474(3)

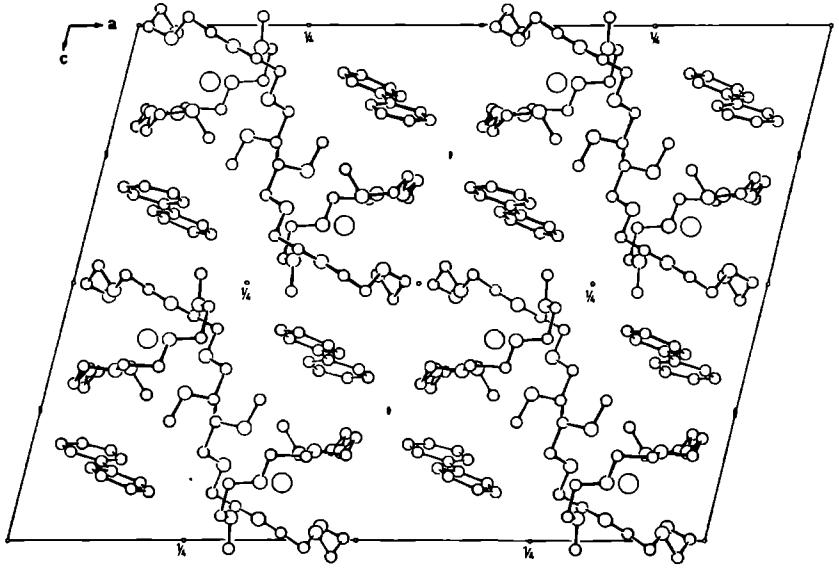


Figure 1. The ac projection of the unit cell of $Rb^{+}\{CH_3O(CH_2CH_2O)_4CH_3\}_2\ C_{12}H_{10}^{-}$. *The symmetry elements are indicated in the picture. The circles of decreasing size are respectively: Rb, O and C.*

tional R-value is 0.054 for all 3573 observed reflections[11]. The atomic scattering factors used were those for Rb^+ (corrected for anomalous scattering $\Delta f'$) O, C and H as given in the International Tables for X-ray Crystallography[12].

Results and Discussion

Crystal structure

The molecular structure (Figure 1) can be described as a solvent seperated ion pair, consisting of a rubidium cation coordinated to two tetraglyme molecules and a biphenyl anion. The ten coordinating oxygen atoms of the two glymes, lie on a sphere around the rubidium ion as can be seen in the stereoscopic picture (Figure 2). The average rubidium-oxygen bond length of 3.01(3)Å resembles the average rubidium-oxygen bond length in dirubidium cyclooctatetraene diglyme (3.02(15)Å)[5]. As can be seen from the difference of the average rubidium-oxygen distance and the Van der Waals radius of oxygen (1.40Å)[14] the Van der Waals radius of the rubidium ion (1.60Å) is slightly larger than its ionic radius (1.47Å)[14]. Bond lengths and bond angles of the biphenyl anion are shown in Figure 3.

Although the phenyl rings are planar within experimental error (maximum deviation of the least squares plane is 0.008Å) the dihedral angle between the two rings turns out to be 9.4°.

Table 3. Bond distances (in Å) and angles (in degrees) for the rubidium tetraglyme part of $Rb^+\{CH_3O(CH_2CH_2O)_4CH_3\}_2C_{12}H_{10}^-$ *(only minimal, maximal and average values are given for corresponding bonds and angles).*

	smallest value	largest value	average value
Rb-O	2.928	3.154	3.015
Tetraglyme molecules			
C-O	1.410	1.452	1.430
C-C	1.484	1.517	1.501
C-O-C	110.1	113.7	111.7
O-C-C	107.3	114.7	109.1

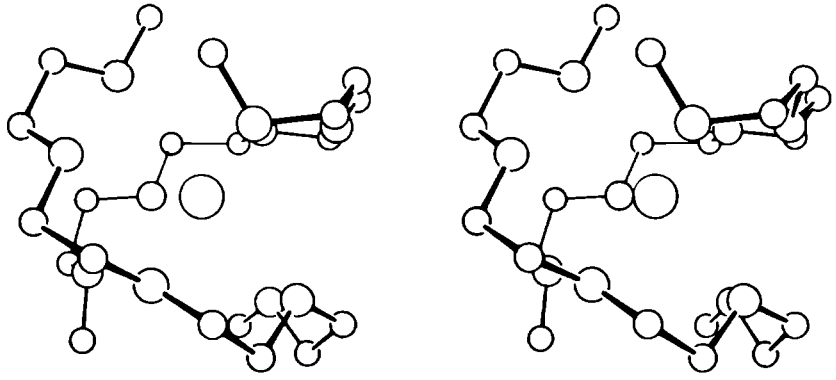


Figure 2. A stereoscopic view of the solvatation of rubidium by two tetraglyme molecules.

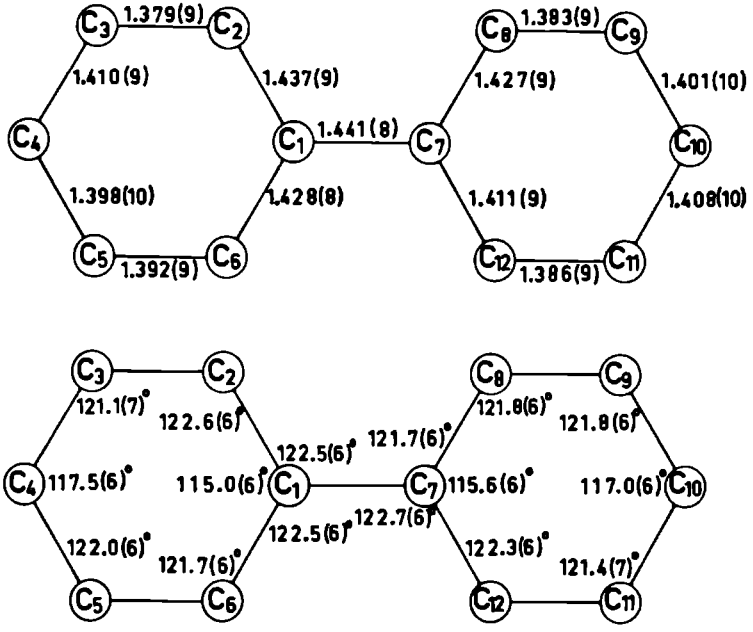


Figure 3. Bond distances (in Å) and angles (in degrees) in the biphenyl anion with standard deviation in parentheses.

Acknowledgements

One of the authors (J.J.M) enjoyed the helpful discussions with Dr. J.M. Trooster about the making of the sweep unit for the susceptibility measurements. He is also very thankful to Mr. C.J. Beers for his skillful assistance in carrying out the foner susceptibility experiments. Part of this work (Th.v.d.H.) has been carried out under the auspices of FOMRE with financial support of the Netherlands Organization for the Advancement of Pure Research (ZWO).

References

(1) M.Szwarc, "Ion and Ion Pairs in Organic Reactions", Vol. 1, Wiley-Interscience, New York, 1972; E.S. Petrov, M.I. Terekhova, and A.I. Shatenshtein, Russ. Chem. Rev., 42, 713 (1973); N.L. Holy, Chem. Rev., 74, 243 (1974).

(2) G.W. Canters, A.A.K. Klaassen, and E. de Boer, J. Phys. Chem., 74, 3299 (1970).

(3) (a) J.J. Brooks and G.D. Stucky, J. Amer. Chem. Soc., 94, 7333 (1972).
(b) J.J. Brooks, W. Rhine, and G.D. Stucky, ibid., 94, 7339 (1972).
(c) J.J. Brooks, W. Rhine, and G.D. Stucky, ibid., 94, 7346 (1972).

(4) S.Z. Goldberg, K.N. Raymond, C.A. Harmon, and D.H. Templeton, ibid., 96, 1348 (1974).

(5) (a) J.H. Noordik, Th.E.M. van den Hark, J.J. Mooij, and A.A.K. Klaassen, Acta Crystallogr., B30, 833 (1974).
(b) J.H. Noordik, H.M.L. Degens, and J.J. Mooij, ibid., in press.

(6) R.H. Cox, L.W. Harrison, and W.K. Austin, J. Phys. Chem., 77, 200 (1973).

(7) P.W. Anderson and P.R. Weiss, Rev. Mod. Physics, 25, 269 (1953).

(8) J.H. van Vleck, Phys. Rev., 74, 1168 (1948).

(9) W.R. Busing and H.A. Levy, Acta Crystallogr., 10, 180 (1957).

(10) R.O. Gould, Th.E.M. van den Hark, and P.T. Beurskens, Acta Crystallogr., to be published.

(11) Observed and calculated structure factors are available from the authors.

(12) International Tables for X-ray Crystallography, Vol. 3, Kynoch Press, Birmingham, 1962.

(13) G.T. Pott, Molionic Lattices (thesis), Groningen, 1966, pp 19-24.

(14) "Handbook of Chemistry and Physics", 49th edition, The Chemical Rubber Co., Ohio, 1968.

CHAPTER 16

PROCEDURE DIRDIF.D

A description of the procedure DIRDIF.D is given in this chapter.

Submitted for publication in Acta Crystallographica.

THE APPLICATION OF DIRECT METHODS TO NON-CENTROSYMMETRIC STRUCTURES CONTAINING HEAVY ATOMS

Th.E.M. van den Hark, Peter Prick and Paul T. Beurskens
Crystallography Laboratory, University of Nijmegen, Toernooiveld,
Nijmegen, The Netherlands

Abstract.

Direct methods are applied to the difference structure factors for a structure containing one or more heavy atoms in known positions. The present procedure is initiated by substracting the known heavy-atom contribution from the observed structure factor (assuming that the observed and calculated structure factors have the same phase) to obtain the magnitude and phase of the light-atom contribution. The sigma-2 phase relationship (tangent formula) is used to recalculate the phases of the light-atom contributions, and -consequently- to recalculate the magnitude of the light-atom contribution. An iterative procedure is used to optimise the phases and amplitudes before a difference Fourier is calculated. The method is applicable also for the solution of partially known structures.

Introduction.

In paper I (Beurskens and Noordik, 1971) and in paper II (Gould et al., 1975) direct methods were used to solve the phase problem, or to speed up the solution, for centrosymmetric structures containing one or more heavy atoms on known positions. The present paper deals with the non-centric case. A procedure to handle the special case in which the known heavy atoms do not completely fix the origin and/or enantiomorph is under investigation. The general case[x]

[x] A FORTRAN program for the execution of this procedure, DIRDIF.D, is available on request.

is discussed in this paper.

The positions of the known (heavy) atoms determine the structure; the phase problem is solved in principle: the positions of the remaining (light) atoms can be found by standard Patterson or Fourier techniques. The time and effort necessary for finding the light-atom structure is reduced by the present procedure, especially when the heavy atoms are only marginally sufficient to solve the phase problem.

Define for a reflection $\underline{h}$:

$|F_{obs}|$ observed structure factor amplitude, on absolute scale

F_H calculated contribution of the known part of the structure ('heavy' atoms)

ϕ_H phase of F_H

F_L contribution of the remaining part of the structure ('light' atoms), or: the most probable estimate for this contribution.

ϕ_L phase of F_L

F_{obs} a phased value for the observed structure amplitude, defined by: $F_{obs} = F_L + F_H$ (1)
Possible solutions to this equation, for given $|F_{obs}|$, ϕ_L and F_H values will be discussed below.

In conventional procedures, the difference Fourier coefficients:

$$\Delta F_1 = (|F_{obs}| - |F_H|) \exp i\phi_H \quad (2)$$

are calculated and accepted as an estimate for F_L. Only in very favorable circumstances can the complete light atom structure be unambiguously deduced from the Fourier synthesis based on these coefficients. In the present procedure, a sigma-2 (tangent formula) refinement procedure is used to convert input ΔF_1 values to more probable F_L values. This procedure depends on a probability estimate for ΔF_1, relative to the extreme opposite possibility, ΔF_2:

$$\Delta F_2 = (-|F_{obs}| - |F_H|) \exp i\phi_H \quad (3)$$

where now F_{obs} is completely out of phase with F_H.

Note, that $|\Delta F_1| \leq |F_L| \leq |\Delta F_2|$.

Although conventionally ΔF_1 is used as difference Fourier coefficient,

ΔF_2 will be the more probable value for relatively small terms (see below).

In the present procedure use is made of those reflections where ΔF_1 is far more probable than ΔF_2, and ΔF_1 is used as a first estimate of F_L. The application of direct methods then leads to new phases ϕ_L, and the magnitude $|F_L|$ then also is to be recalculated, using (1), which may be written as

$$|F_{obs}| = ||F_L| \exp(i\phi_L) + F_H| \qquad (4)$$

(see figures 1 and 2, to be discussed later).

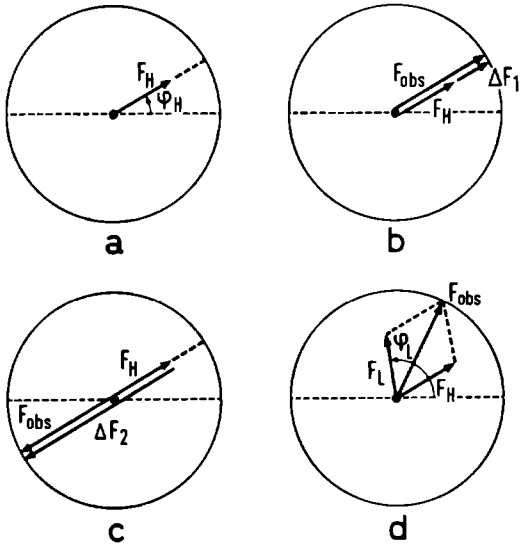


Fig. 1. Definition of ΔF_1 and ΔF_2, and construction of F_L for
case 1: $|F_H| < |F_{obs}|$
a) $|F_{obs}|$-circle, with calculated F_H
b) definition of ΔF_1; F_H and F_{obs} are in phase
c) definition of ΔF_2; F_H and F_{obs} out of phase by 180^o
d) general case; ϕ_L is assumed to be known

Normalisation of the difference structure factors

A modified, two-dimensional Wilson plot, as described in papers I and II, is used to obtain the scale factor K, the overall temperature factor of the heavy atoms B_H, and the overall temperature factor of the light atoms B_L. K is used to bring $|F_{obs}|$ on absolute scale. B_H is used in the calculation of the heavy atom contributions F_H. B_L is used to calculate the normalizing function $g(\underline{h})$:

$$g = (\epsilon_{\underline{h}} \sum_{j=1}^{n_L} f_j^2)^{\frac{1}{2}} \exp -B_L \sin^2\theta/\lambda^2 \qquad (5)$$

where n_L is the number of unknown (light) atoms in the unit cell and ϵ_h is the usual factor to correct for symmetry enhancement for the reflection h. The normalized difference structure factor is defined as:

$$E_L = (F_{obs} - F_H) / g \qquad (6)$$

(for a given reflection h, if the phase of F_{obs} is known).

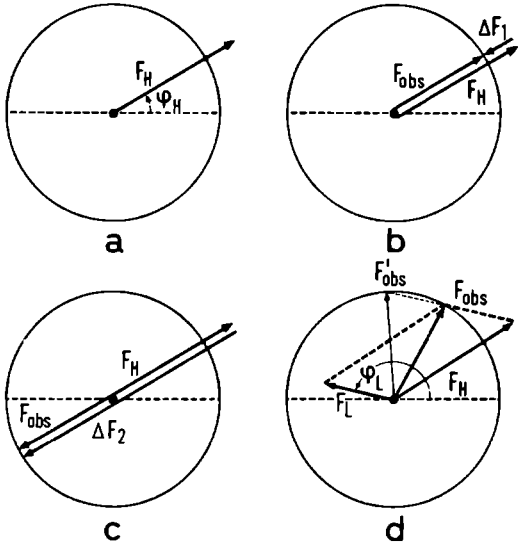


Fig. 2. Definition of ΔF_1 and ΔF_2, and construction of F_L for case 2: $|F_H| > |F_{obs}|$ a) - d) as in figure 1.

The above defined ΔF_1 and ΔF_2 values are brought on the same scale by defining:

$$E_1 = \Delta F_1 / g \qquad \text{and} \qquad E_2 = \Delta F_2 / g \qquad (7)$$

The initially unknown E_L values correspond to the normalized structure factors of a structure consisting of only the unknown (light) atoms. Conventional direct methods may, in principle, be applied to E_L values, in case they are known. For reflections where E_1 is far more probable than E_2, we use E_1 as an initial estimate for E_L and use these to initiate a tangent-formula refinement procedure.

Note: the original $|E_1|$ values do not form a normalized set of structure factors; the average of squares of $|E_1|$ is less than 1. During the refinement of phases, the magnitudes of $|E_L|$ are being increased, on an average, and the distribution of $|E_L|$ values will become more normalized.

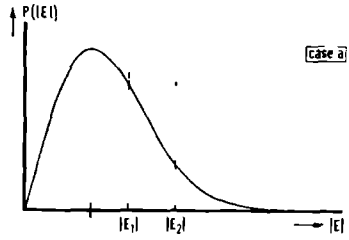
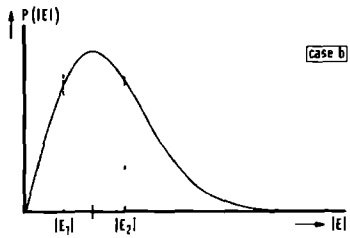


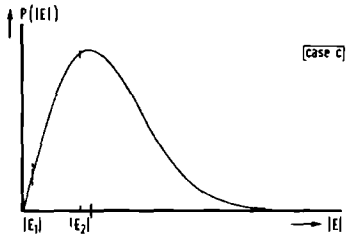
Fig. 3. Classification of reflections depending on $|E_1|$ and $|E_2|$ pairs.

Case a : $|E_2| > |E_1| > 0.7$

Case b : $|E_1| < 0.7 < |E_2|$

Case c : $|E_1| < |E_2| < 0.7$





Probability considerations

The distribution function for acentric reflections is given by

$$P(|E|) = 2|E| \exp -|E|^2 \qquad (8)$$

This function has a maximum at $|E| = \frac{1}{2}\sqrt{2} \approx 0.7$. The occurrence of this maximum forced us to consider three distinct cases (Figure 3).
(Note: special reflections, having a centric distribution, will be treated accordingly; see paper II; we now limit our discussion to general (non-centric) reflections).
For case a: $|E_1| > 0.7$.
Many reflections will have both $|E_1|$ and $|E_2|$ greater than 0.7. The number of reflections that belong to this category depends on the known fraction of the scattering power; usually it is about half the total number of reflections, or less.

This is the most important case, as $|E_1|$ (as well as any possible value for $|E_2|$) is large enough to be of importance in a Fourier synthesis and to be useful in the application of the tangent formula. For the-

se reflections $|E_1|$ is more probable than $|E_2|$, and E_1 may be selected and used as a first-estimate for E_L in the case the probability for $|E_1|$ is significantly greater than the probability for $|E_2|$. Analogous to the centrosymmetric procedure (Paper II) we now use P_1 as a measure for the relative probability of the phase of E_1:

$$P_1 = P(|E_1|) \; / \; \{P(|E_1|) + P(|E_2|)\} \tag{9}$$

and the weight for this reflection is chosen to be:

$$W_1 = (2P_1 - 1)^2 \tag{10}$$

In principle this is not correct; the weight should be based upon the standard deviation σ_1 for the phase of E_1, which can be calculated from eq. (8):

$$\sigma_1^2 = \int_1^2 P(|E|) \{\Delta\phi\}^2 \, d|E| \; / \int_1^2 P(|E|) \, d|E| \tag{11}$$

where the integration limits are $|E_1|$ and $|E_2|$, and:

$$\cos\Delta\phi = (|E_1|.|E_2| - |E|^2) \; / \; (|E_2|.|E| - |E_1|.|E|)$$

Numerical results are given in table 1. (These results should also be used in the treatment of the weaker reflections).

Table 1. Standard deviations (^{o}) of ϕ_1 for pairs of $|E_1|$ and $|E_2|$ values, calculated from eq. (11).

$\|E_2\|$ =	0.3	0.6	0.8	1.0	1.2	1.4	1.7	2.0	2.5	3.0	4.0
$\|E_1\|$ = 0.3	104	109	111	111	110	108	104	99	92	88	82
$\|E_1\|$ = 0.6		104	101	102	101	99	94	89	81	75	69
$\|E_1\|$ = 0.8			104	99	98	96	91	85	76	70	63
$\|E_1\|$ = 1.0				104	97	95	90	83	73	66	58
$\|E_1\|$ = 1.2					104	96	90	83	71	63	54
$\|E_1\|$ = 1.4						104	92	85	71	61	52
$\|E_1\|$ = 1.7							104	90	74	61	49
$\|E_1\|$ = 2.0								104	80	63	47
$\|E_1\|$ = 2.5									104	75	47
$\|E_1\|$ = 3.0										104	51
$\|E_1\|$ = 4.0											104

In practice, however, eq. (9) gives acceptable results for the first input phases for the tangent refinement procedure, and the weights for the following cycles of the refinement will greatly be determined by the results of the tangent formula.

For case b, $|E_1| < 0.7 < |E_2|$, and for case c, $|E_2| < 0.7$, the value of $|E_1|$ is <u>not</u> the most probable value. The tangent formula cannot effect the phase of such a reflection; nevertheless E_1 or E_2 of the reflection may be used as Fourier coefficient if justified by its relative probability (eq. 9) or, better, its standard deviation (eq. 11). In case $|E_1|$ is very small (equal or less its estimated error) the reflection is excluded from further calculations.

Tangent refinement procedure

The case a - reflections may enter into the tangent refinement procedure. As input to the tangent formula we use those reflections where $|E_1|$ exceeds a given minimum value greater than 0.7 (say: $|E_1| > 1.2$). We then use this formula to calculate phases for all reflections where $|E_1|$ exceeds another minimum greater than 0.7 (say: $|E_1| > 0.9$).

The tangent formula may be given as:

$$\phi_L(t) = \text{phase of } (E_L)_{\underline{h}} \approx \text{phase of } \Sigma_2 \qquad (12)$$

with $\Sigma_2 = \sum_{\underline{k}} W_{\underline{k}} \; W_{\underline{h}-\underline{k}} \; E_{\underline{k}} \; E_{\underline{h}-\underline{k}}$

and $E_{\underline{k}}$ is the most probable E_L value for the reflection $\underline{k}$. Analogous to the centrosymmetric formula, and to the formulae (9) and (10), we use the following simple expression for the relative probability and the corresponding weight:

$$P_t = \tfrac{1}{2} + \tfrac{1}{2} \tanh \sum_j Z_j^3 \left(\sum_j Z_j^2 \right)^{-3/2} |E_{\underline{h}}| \; \Sigma_2 \qquad (13)$$

$$W_t = (2P_t - 1)^2 \qquad (14)$$

where Z_j is the number of electrons for the jth atom. {For the weighted tangent formula, see Germain et al., 1971}

In the first cycle we only have E_1 values (with phases ϕ_1) as first estimates for E_L for the reflections $\underline{k}$ and $(\underline{h}-\underline{k})$. Whether or not the output phases ϕ_L are accepted depends on the corresponding weights. If $W_t > W_1$ then the new phase ϕ_L is accepted with weight W_t. If $W_t < W_1$ then the calculated ϕ_L value is only partially accepted, in case

$|\phi_L(t) - \phi_1|$ is less than 90^o:

$$\phi_L(new) = \phi_1 + (W_t / W_1) \{\phi_L(t) - \phi_1\} \quad (15)$$

In addition, there is a limitation on $\{\phi_L(t) - \phi_1\}$ for reflections of category a2, see below, and ϕ_L may be reset accordingly. The new ϕ_L value is used to calculate a new value for E_L, eq. (4), which may be used as input for the next tangent refinement cycle.

Note: As a consequence of the definitions for W_1 and W_t (given in eq. 10 and 14, which on an average result in a slightly overestimated W_1 relative to W_t) the original phases ϕ_1 still play an important role in the second refinement cycle. Using a weighting scheme based upon eq. (11) calls for the replacement of the conventional scheme (14) by a scheme that is based upon the standard deviations associated with the tangent formula.

Description of the procedure

We have assumed that the heavy atom part of the structure is known, and that its contribution to the structure factors are calculated. The normalisation procedure, described above, will then lead to values for $|F_{obs}|$, F_H, ΔF_1, ΔF_2, E_1, E_2, P_1 and W_1.

The distinction in several cases, as given above, leads to the following categories of reflections: a1, a2, b1, b2, c1 and c2. A detailed description for each of these categories is given below.

At the end the final E_L values are transformed back to F_L values and used in a weighted Fourier synthesis.

Category a1. For reflections with $|F_H| < |F_{obs}|$ we have (see fig. 1)

$$\phi_1 = \phi_H = \phi_2 + 180^o \quad (16)$$

The strongest of these form the category a1. This is the most important set of reflections. In principle all of these reflections may be treated alike; in practice, however, a considerable amount of computer time is saved by limiting the number of reflections that enter into the tangent refinement procedure. These reflections will have, after four to six cycles of tangent refinement, a calculated E_L value. If this calculated value is unreliable, then the original E_1 value will be used according to formula (7).

The remaining reflections are those with $0.7 < |E_1| < E_{min}$ (=say:

0.9); for these reflections the E_1 value will be used, with its proper weight W_1.

Category a2. For reflections with $|F_H| > |F_{obs}|$ we have (see fig. 2):

$$\phi_1 = \phi_2 = \phi_H + 180^\circ \qquad (17)$$

As can be seen from the diagram in fig. 2d, there are two possible $|E_L|$ values for a given phase ϕ_L; the smallest $|E_L|$ value is the most probable one and, naturally, this is the value to be used in our procedure. For a given $|F_{obs}|$ and F_H there is a restriction on the possible phase values ϕ_L, see fig. 4.

Let us define: $\Delta\phi = \phi_L - \phi_1$ (18)

that is: $\Delta\phi$ is the correction to the original ϕ_1 value; we hope to find $\Delta\phi$ by the weighted tangent procedure. The maximum value for $|\Delta\phi|$ is given by:

$$\sin|\Delta\phi_{max}| = |F_{obs}| \; / \; |F_H| \qquad (19)$$

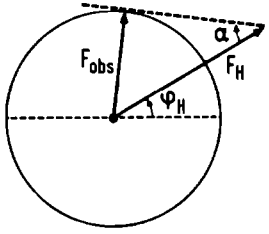


Fig. 4. Definition of the limiting angle for reflections with: $|F_H| > |F_{obs}|$; $\alpha = \Delta\phi_{max}$ *(eq. 19).*

For $|E_1| > 0.7$ and (most likely) $|E_2| < 4.0$ we have $\Delta\phi_{max} = 44.6^\circ$ as the largest possible value for $\Delta\phi_{max}$; so the few reflections in this category have well determined phases and will therefore be given unit weight ($W_1 = 1$) in all calculations. If the $\Delta\phi$, calculated by the tangent formula (12), exceeds $\Delta\phi_{max}$ for the given reflection, then $\Delta\phi_{max}$ is substituted for $\Delta\phi$. In the last cycle, however, the calculated $\Delta\phi$ is accepted to allow for possible experimental errors in $|F_{obs}|$ and model-errors in F_H.

Category b1. Because of the low $|E_1|$ value the tangent formula may lead to incorrect results, and a change in phase, leading to larger $|E_L|$ values cannot be trusted; so the tangent formula is not used at all. Although the reflections in this category have rather low $|E_1|$ values,

this category is not unimportant because of the large number of reflections. E_1 values will be used, with its proper weight W_1, only in case $W_1 > W_2$. Reflections with $W_1 < W_2$ are rejected for several reasons (see also category c1).

Category b2. The reflections belonging to this category have reliable phases ϕ_1 ($W_1 = 1$, see category a2) and are therefore useful for the Fourier synthesis; ϕ_L is taken as ϕ_1, and the absolute value $|E_L|$ is taken as its expectation value:

$$\int_1^2 |E|\, P(|E|)\, d|E| \Big/ \int_1^2 P(|E|)\, d|E| \qquad (20)$$

Category c1. For these reflections E_2 is a more probable estimate for E_L than E_1 (in contrast to the conventional definition of difference Fourier coefficients). It is certainly not useful to put these reflections as E_1 into a Fourier synthesis. To use E_2 as Fourier coefficients on the other hand, may easily lead to an increased Fourier-noice level because of the uncertainties in $|F_{obs}|$ and F_H. At present we reject these reflections, but we will change our computer program as soon as our experiments show the usefulness of these reflections.

Category c2. This small set of reflections is treated as category b2.

Examples

The procedure has been used successfully in the structure analysis of $(HgI_2)_2\ CH_2\ (S_2CNC_4H_{10})_2$, and of two modifications of heptahelicene, $C_{30}H_{18}$. In the former structure one HgI_2 unit was found from the Patterson synthesis, expecting the space group to be Cc; application of the program DIRDIF.D revealed the unexpected presence of a second HgI_2 unit (Beurskens et al., 1976a). The application of DIRDIF.D on the heptahelicene structures, where a fragment of the structure was found by direct methods, will be described in a consecutive communication (Beurskens et al., 1976b).

Some numerical results for a test structure will be given in tables 2-4: 4,4 dichloro-2a-aza-A-homocholestan-3-on, $C_{27}H_{45}NOCl_2$, space group $P2_12_12_1$, Z=4 (Mootz and Berking, 1970). The published parameters of the two chlorine atoms were used as heavy atoms; the Fourier-coefficients produced by DIRDIF.D gave an electron-density map that revealed the molecule much better than the normal difference Fourier synthesis. The improvement of the phases is shown in tables 3 and 4.

Table 2. Number of reflections of the test structure in each of the six categories.

Category	Number of reflections	
a1	458	(360 reflections with $\|E_l\| > 0.9$)
a2	62	(18 reflections with $\|E_l\| > 0.9$)
b1	613	
b2	364	
c1	399	
c2	45	

Table 3. Average deviations of refined phases for the test structure.

Ranges in W_l \ Ranges in phase deviation before tangent refinement[x]		0-30°	30-60°	60-90°	90-180°
Category a1:					
0.0 - 0.3	After[xx]	14°(6)	10°(6)	26°(9)	23°(15)
0.3 - 0.6		14 (6)	16 (8)	20 (5)	33 (6)
0.6 - 0.99		12 (23)	22 (21)	29 (13)	6 (4)
> 0.99		15 (51)	17 (39)	12 (17)	12 (6)
Category a2:					
$W_l = 1.0$		15 (13)	47 (3)	3 (1)	30 (1)

[x] ranges in $|\phi_l - \phi_{calc}|$, see table 4.

[xx] average in $|\phi_L - \phi_{calc}|$, after the tangent refinement for the reflections in the given ranges (the number of reflections is given between parentheses).

Table 4. Results of the tangent refinement procedure for the test structure, for reflections of the categories a1 and a2 with $|E_1| > 0.9$.

Ranges in W_1 \ Ranges in phase deviation[x]		0-30°	30-60°	60-90°	90-180°	Average phase deviation	Number of reflections
Category a1:							
0.0 - 0.3	Before[xx]	6	6	9	15	85°	36
	After[xx]	29	4	2	1	20°	
0.3 - 0.6	Before	6	8	5	6	62°	25
	After	22	2	0	1	21°	
0.6 - 0.99	Before	23	21	13	4	45°	61
	After	46	12	3	0	19°	
> 0.99	Before	51	39	17	6	39°	113
	After	98	14	1	0	15°	
Category a2:							
$W_1 = 1.0$	Before	13	3	1	1	28°	18
	After	13	5	0	0	20°	
Projection reflections:							
0.0 - 1.0	Before	94	-	-	31	-	125
	After	121	-	-	4	-	

x Phases are compared with the true phases ϕ_{calc}, calculated from the final light atom structure. Before applying the tangent formula, the 'phase deviation' of a reflection is defined as $|\phi_1 - \phi_{calc}|$. After the tangent refinement the 'phase deviation' is defined as $|\phi_L - \phi_{calc}|$.

xx Tabulated are the number of reflections belonging to the specified ranges in W_1 and 'phase deviation', before and after the tangent refinement

Conclusions

The procedure described above proved to be useful in routine structure analysis of heavy atom compounds, as well as for partially known equal atom structures. The special case, where the origin and/or enantiomorph is not completely fixed by the known part of the structure is presently being studied.

Acknowledgement.

One of us (Th.E.M. van den Hark) acknowledges support of the Dutch Foundation for Pure Research, ZWO/FOMRE.

References

Beurskens, P.T. and Noordik, J.H. (1971). Acta Cryst. A27, 187.

Beurskens, Paul T., Enckevort, W.J.P., Menger, E.M. and Bosman, W.P. (1976a). To be published.

Beurskens, Paul T., van den Hark, Th.E.M. and Beurskens, Gezina (1976b). Acta Cryst. A32, 000.

Germain, G. Main, P. and Woolfson, M.M. (1971). Acta Cryst. A27, 368.

Gould, R.O., van den Hark, Th.E.M. and Beurskens, Paul T. (1975). Acta Cryst. A31, 813.

Mootz, Dietrich and Berking, Bernhard (1970). Acta Cryst. B26, 1362.

CHAPTER 17

'HEPTA.C'

17.1 Crystal structure of 'HEPTA.C'

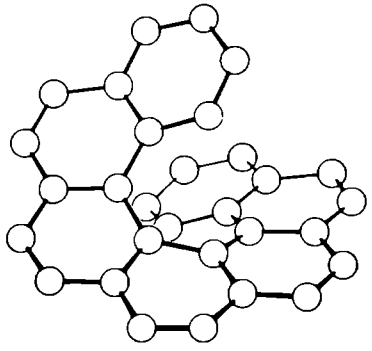
We will reproduce in this section the paper published in Cryst. Struct. Comm. (1976) 5, 000

HEPTAHELICENE, $C_{30}H_{18}$ (2^{nd} modification)

Th.E.M. van den Hark and Paul T. Beurskens

Crystallography Laboratory, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands.

Preliminary information. The present compound was prepared by Laarhoven et al. (1970). Heptahelicene crystallizes in two modifications; the crystal structure of the first modification in which heptahelicene crystallizes in the space group $P2_1$ with two molecules in the asymmetric unit has recently been published by Beurskens et al. (1976). We now report the crystal structure determination of the second modification, where heptahelicene crystallizes in the space group $P2_1/c$. This structure has also been solved by Germain et al. (1975).



Crystal data. (from single crystal diffractometry; $\lambda MoK\alpha$ = 0.71069Å) Monoclinic, a = 8.112(7), b = 11.504(13), c = 21.584(21)Å, and β = 106.23(7)°. Space group $P2_1/c$, Z = 4, V = 1934Å^3, D_x = 1.30g/cm^3, F.W. = 378.5.

Intensity data, structure determination and refinement. Intensity data were collected with Zr-filtered MoKα radiation on a NONIUS-CAD3-diffracto-

meter ($\theta - 2\theta$ scan). Of the 1757 attainable symmetry independent reflections up to $\sin\theta/\lambda = 0.48\text{Å}^{-1}$, 631 reflections were 'observed' with intensities greater than five standard deviations, based on counting statistics. (The weaker reflections appeared to be measured very badly, due to instrumental troubles).

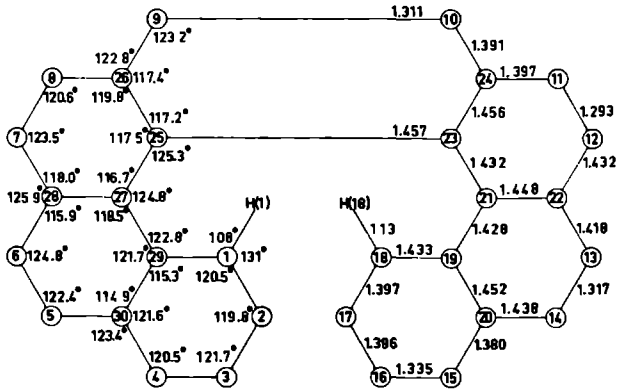
Routine applications of direct methods failed, probably because of the relatively small number of reflections. The structure was then solved by unconventional direct methods techniques. Application of the B3,0 formula (van den Hark et al.,1975) resulted in a sixteenfold solution for the phase problem. This number was reduced to four by application of sigma-3 and pair relationships (Beurskens et al.,1975). One of the four solutions revealed a clearly recognisable molecule. Refinement of this structure with least-squares techniques, however, was unsuccessful. After reduction of the space group symmetry to P1, we used one known heptahelicene molecule to phase the difference structure factors; the remaining three molecules were found with the program DIRDIF.D (van den Hark et al., 1976). It proved that the original heptahelicene molecule was shifted perpendicular to the c-glide plane over a distance of 0.71Å. Atomic coordinates and isotropic thermal parameters of all carbon atoms were refined by means of full-matrix least-squares refinement techniques, minimizing the function $\Sigma\, w\, \{|F_o| - k|F_c|\}^2$, with weights $w = \{\sigma_c^2(F_o) + (0.05|F_o|)^2\}^{-1}$ ($\sigma_c(F_o)$ from counting statistics), using only the observed reflections. All hydrogen atoms, except H(1) and H(18), were introduced at calculated positions with a C-H bond length of 1.084Å. From a difference Fourier map, calculated from reflections with $\sin\theta/\lambda < 0.3\text{Å}^{-1}$, the positions of H(1) and H(18) were found. All hydrogen atoms were included in the structure factor calculations with an isotropic temperature factor of $B = 5.0\text{Å}^2$, but were not refined. The lack of data did not permit anisotropic refinement. The final conventional R-factor is 0.095 for the 631 observed reflections.

Comments. The molecule possesses a pseudo(local) two-fold symmetry axis; the deviations from true symmetry (with respect to bond distances and angles) are in the order of the standard deviations. Therefore, corresponding bond distances and angles have been averaged (e.s.d. of the averages: 0.018Å and 1.2°).

Atomic coordinates.[x]

Atom	x	y	z	Atom	x	y	z
e.s.d.	0.003	0.002	0.0010	e.s.d.	0.003	0.002	0.0010
C(1)	0.453	0.210	0.0329	C(18)	0.205	0.274	0.1543
C(2)	0.362	0.163	-0.0271	C(17)	0.038	0.316	0.1330
C(3)	0.232	0.228	-0.0689	C(16)	-0.089	0.250	0.0908
C(4)	0.190	0.333	-0.0543	C(15)	-0.057	0.140	0.0755
C(5)	0.249	0.503	0.0190	C(14)	0.149	-0.025	0.0879
C(6)	0.354	0.553	0.0715	C(13)	0.297	-0.072	0.1145
C(7)	0.593	0.551	0.1747	C(12)	0.607	-0.054	0.1859
C(8)	0.713	0.493	0.2170	C(11)	0.731	0.004	0.2208
C(9)	0.844	0.308	0.2615	C(10)	0.848	0.194	0.2593
C(25)	0.591	0.310	0.1686	C(23)	0.571	0.186	0.1791
C(26)	0.721	0.373	0.2182	C(24)	0.717	0.126	0.2200
C(27)	0.491	0.372	0.1143	C(21)	0.418	0.117	0.1538
C(28)	0.477	0.495	0.1208	C(22)	0.444	-0.008	0.1500
C(29)	0.404	0.320	0.0524	C(19)	0.250	0.162	0.1348
C(30)	0.273	0.382	0.0033	C(20)	0.109	0.093	0.0970
H(1)	0.575	0.180	0.070	H(18)	0.319	0.307	0.194

[x]Two atoms on one line are related by the molecular symmetry axis.



Some individual contact distances in the molecule are:

C(1) - C(21)	2.906(22) Å	C(18) - C(27)	2.922(22) Å
C(1) - C(23)	3.043	C(18) - C(25)	3.092
H(1) - C(23)	2.36	H(18) - C(25)	2.42

Torsional angles

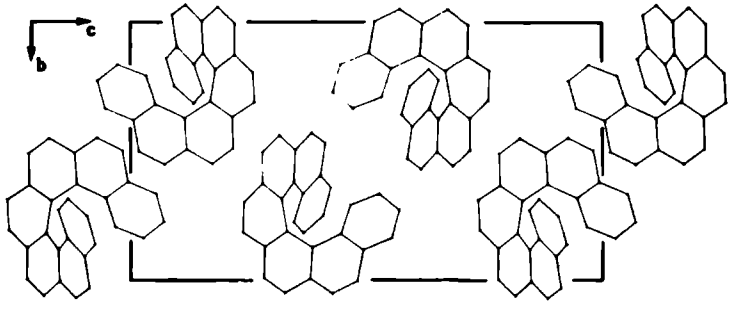
H(1) -	C(1) -	C(29) -	C(27)	6.7°
C(1) -	29 -	27 -	25	20.8
29 -	27 -	25 -	23	23.7
27 -	25 -	23 -	21	25.2
25 -	23 -	21 -	19	23.8
23 -	21 -	19 -	18	18.0
21 -	19 -	18 -	H(18)	11.1

Interplanar angles for consecutive C_6-rings.

A,B	10.9°
B,C	11.3
C,D	11.7
D,E	13.0
E,F	11.9
F,G	9.9
(A,G	32.3)

The bond lengths and angles of the molecule are the same (within the accuracy of the structure determination) as the bond lengths and angles of both molecules of the first modification (Beurskens et al., 1976). The slight differences in the overall geometry of the molecules are caused by packing differences for the three independent molecules.

The averaged molecule is given in the table in cartesian coordinates (Å): the origin is at the center of gravity; the molecular symmetry axis is along the Z axis; the center of gravity of the atoms C(11) - C(14) and C(19) - C(24) is in the YZ-plane. This molecule is the average of the three molecules (giving double weight to the two molecules of the first modification) idealized with respect to the molecular symmetry axis.



Projection along the a-axis.

(That is: a coordinate given in this table is the average of six independent measurements; the e.s.d. in the averaged atomic positions is 0.003 Å).

Cartesian coordinates for the averaged molecule (see text).

	X	Y	Z
C(1)	1.913 Å	-0.086 Å	-0.701 Å
C(2)	2.612	0.265	-1.832
C(3)	2.435	-0.444	-3.024
C(4)	1.617	-1.523	-3.045
C(5)	0.210	-3.170	-1.880
C(6)	-0.301	-3.621	-0.739
C(7)	-0.749	-3.417	1.699
C(8)	-0.686	-2.719	2.834
C(9)	-0.197	-0.653	4.016
C(25)	-0.013	-0.723	1.576
C(26)	-0.304	-1.362	2.821
C(27)	0.224	-1.554	0.436
C(28)	-0.294	-2.883	0.467
C(29)	0.998	-1.163	-0.718
C(30)	0.923	-1.937	-1.910

Acknowledgements. The authors wish to thank Dr. W.H. Laarhoven for suggesting this research and for providing the crystals and Mr. J.M.M. Smits for performing the X-ray data collection. One of us (Th.E.M. v.d. Hark) acknowledges support of the Netherlands Organization for Pure Research (ZWO/FOMRE).

References.

Beurskens, P.T. and van den Hark, Th.E.M. (1975). Acta Cryst. A31, S14.

Beurskens, Paul T., Beurskens, Gezina and van den Hark, Th.E.M. (1976). Cryst. Struct. Comm. 5,

Germain, G. and collaborators (1975). Private communication.

Laarhoven, W.H., Cuppen, Th.J.H.M. and Nivard, R.J.F. (1970). Tetrahedron 26, 4865.

Van den Hark, Th.E.M. and Beurskens, P.T. (1975). Acta Cryst. A31, S17.

Van den Hark, Th.E.M., Prick, Peter and Beurskens, Paul T. (1976). Submitted for publication.

17.2 Comments on the crystal structure determination of 'HEPTA.C'

The crystal structure determination of a compound of the complexity of 'HEPTA.C', in a centrosymmetric space group, by direct methods gives, in general, little trouble. The crystal structure determination of 'HEPTA.C', however, was certainly not straightforward, with the available reflection data.

After completing the structure we checked the performance of the sign determination. In this section we will present some of our observations and will try to relate the troubles encountered during the structure determination to the crystal structure.

The Fourier synthesis which led to the identification of a heptahelicene molecule will be discussed here. This Fourier map was calculated from the solution that was most consistent with respect to the sigma-2 relationship and which belonged to the top four solutions indicated by the sigma-3 and pair relationships. Moreover, this map showed the most reasonable distribution of peak heights among the sixteen different Fourier maps. The strongest 41 peaks from this Fourier map along with five weaker peaks are displayed in fig. 17.1. Two heptahelicene molecules can be constructed from these peaks. In each of them one atom is missing. The two molecules have the strongest twelve peaks in common.

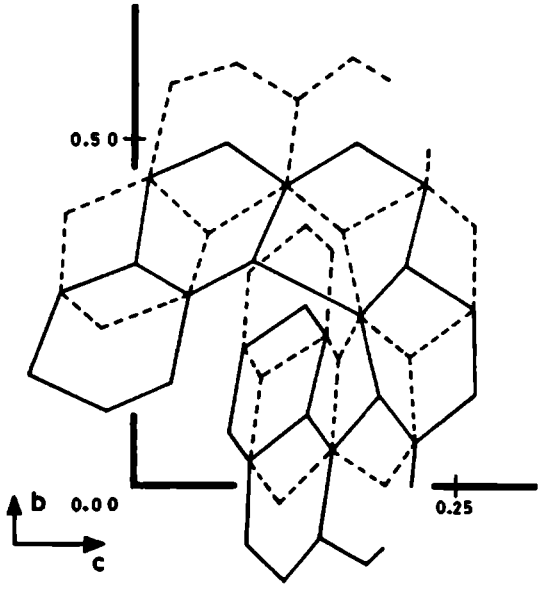


Figure 17.1
The two heptahelicene molecules, as they were found in the Fourier synthesis (projection along the a^* axis).

The two molecules are shifted with respect to each other over a distance of 1.4 Å (≡ one C-C bond distance) along the b-axis. Both molecules proved to be in the correct orientation but at an incorrect position in the unit cell. One of the molecules was used in the program DIRDIF.D to find its correct location in the cell (see chapter 18). The true position of the molecule proved to be in between the molecules displayed in fig. 17.1. Note that the displacement is perpendicular to the glide plane and that symmetry related molecules are also displaced parallel to the b-axis.

The fact that two molecules are found, both incorrectly placed in the unit cell, is related to the observation that the reflections, used for the calculation of the Fourier map, can be divided in two classes:

a) reflections with k-indices smaller than or equal to four; almost all of these reflections had correct signs.
b) reflections with k-indices greater than four; almost all of these reflections were incorrectly signed.

The sign determination was done with the sign correlation procedure applied to the results of the B3,0 formula. We calculated AB_{hk} + C values for all 1409 triples[x], generated from all reflections with $|\overline{E}| > 1.5$ (exclusive sigma-1 type sigma-2 interactions; see section 7.2). The sign determination was done very carefully: in the initial stage only the top results of the B3,0 calculations were used (table 17.2); later on more and more triples were added. Despite this fact, we did not obtain the correct solution.

What caused the trouble? Already in the first step of the sign correlation procedure, some reflections were accepted with incorrect signs. These reflections led very quickly to more incorrect signs. Two facts are important in connection with these incorrect sign assignments:

a. A minus sign was found for the reflection 060, and accepted; it proved to be wrong afterwards. This sign indication came from the sigma-1 relationship and was affirmed by the sigma-2 relationship and by B3,0 calculations for several sigma-1 type sigma-2 interactions.

---

[x]Usually we calculate fewer triple invariants. Because a first attempt to solve this structure was unsuccessful, we decided to calculate this number of triple invariants.

b. A relatively large number of triples of the types:

$$S_{h_1 3 1_1} \cdot S_{h_2 2 1_2} \cdot S_{-h_1-h_2\ \bar{5}\ -1_1-1_2} \qquad (17.1a)$$

and

$$S_{h_1 3 1_1} \cdot S_{h_2 3 1_2} \cdot S_{-h_1-h_2\ \bar{6}\ -1_1-1_2} \qquad (17.1b)$$

for arbitrary h_1 and h_2 values and for even 1_1 and 1_2 values, are incorrect (table 17.1).

Table 17.1 Number of correct and incorrect triples of the type (17.1) in the list of 1409 triples.

	Triples (17.1a)	Triples (17.1b)
Correct triples	12	4
Incorrect triples	38	10

Almost all triples of table 17.1 that proved to be incorrect, were generated from low-order reflections; in all correct triples high-order reflections were also involved.

The results of the B3,0 formula are presented in table 17.2; the results for the triples of the type (17.1) are given separately from the results for the remaining triples. Almost all incorrect triples in the top of the B3,0 results (i.e. with high W_{hk} values) are of the type (17.1).

Table 17.2 Results of the B3,0 formula for 'HEPTA.C'.

W_{hk} range	Number of triples (17.1)		Number of remaining triples	
	Total	Incorrect	Total	Incorrect
> 5.0	7	7	159	0
3.0 - 5.0	18	18	181	5
2.0 - 3.0	13	9	227	14
1.0 - 2.0	12	9	385	49
0.0 - 1.0	9	4	197	40
0.0	5	1	196	93

The incorrect sign of reflection 060. In order to explain this, the atoms of the carbon skeleton of the structure will be divided in two groups (fig. 17.2 and table 17.3). The contribution to the normalized structure factor for a reflection for each of the two groups can be calculated from eq. (2.1), where the summation, $\sum_{j=1}^{N}$, now is over the carbon atoms of each of the two groups, including symmetry-related atoms. The contribution of the atoms of group (1) is denoted by $E_{\underline{h}}^{(1)}$ and of group (2) by $E_{\underline{h}}^{(2)}$; then $E_{\underline{h}\,calc} \approx E_{\underline{h}}^{(1)} + E_{\underline{h}}^{(2)}$ (hydrogen atoms will be neglected).

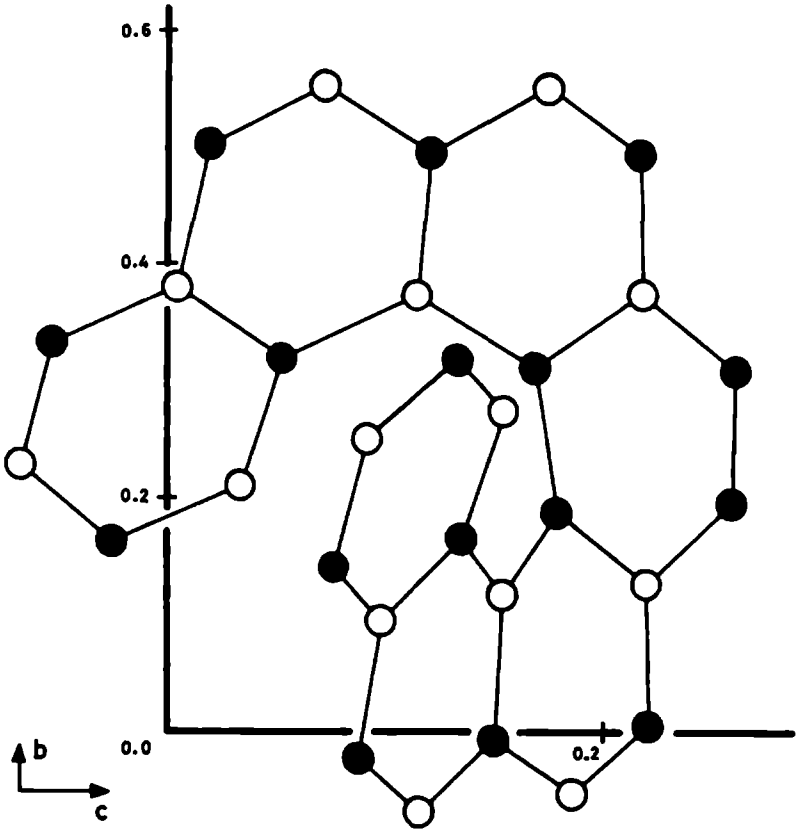


Figure 17.2 The two groups of carbon atoms in the crystal structure of 'HEPTA.C'. Atoms denoted by ● belong to group (1), atoms denoted by ○ to group (2).

Table 17.3 The two groups of carbon atoms in the crystal structure of 'HEPTA.C'. The atomic numbering is as in the preceding section.

Group (1)	Group (2)
C(2) , C(4) , C(5) , C(8) ,	C(1) , C(3) , C(6) , C(7) ,
C(9) , C(10), C(11), C(14),	C(12), C(13), C(16), C(18),
C(15), C(17), C(19), C(22),	C(20), C(21), C(24), C(26),
C(23), C(25), C(28), C(29)	C(27), C(30)

The atoms of group (1) have a very strong contribution to the normalized structure factor for the reflection 060 (82% of its maximum possible value); those of group (2) have also a strong (but smaller) contribution. The contributions have different signs. All h31 reflections with even l-indices, on the contrary, have a weak contribution from the atoms of group (1); they are almost completely determined by the atoms of group (2). Reflections (h31) with odd l-indices also have a contribution from atoms of group (1); high-order reflections frequently have a strong contribution. In table 17.4a are the calculations for the strongest h31 reflections listed.

The sigma-1 relationship will indicate a correct sign for $E_{\underline{h}}^{(1)}$ and $E_{\underline{h}}^{(2)}$ for the reflection 060, when both groups of atoms, (1) and (2), are considered separately. For example, the reflection 236 has an $|E_{\underline{h}}^{(1)}|$ value much smaller than one and an $|E_{\underline{h}}^{(2)}|$ value much greater than one. This results, using the space group symmetry, in a positive and a negative sign, respectively, for $E_{\underline{h}}^{(1)}$ and $E_{\underline{h}}^{(2)}$ for the reflection 060. For the complete structure, an incorrect sign for the reflection 060 is obtained, because the atoms of group (1) dominate the reflection 060, while the atoms of group (2) dominate the reflection 236.

For the triples of the type (17.1), essentially the same arguments hold. Calculations for several reflections involved in these triples are presented in the tables 17.4a and 17.4b. As was mentioned earlier, triples of the type (17.1) in which high-order reflections were involved proved to be correct. This can be explained by the fact that the systematic pattern in the structure, responsible for the many incorrect triples (17.1), is only approximate. Consequently, if data of more high-order reflections had been available, the structure would have been solved

Table 17.4 Results of the calculation of the contribution to the normalized structure factor for some reflections, for each of the two groups of atoms.

a. The reflection 060 and some of the strongest h3l reflections.

Reflection $\underline{h}$	$E_{\underline{h}}^{(1)}$	$E_{\underline{h}}^{(2)}$	$E_{\underline{h}_{calc}}$	$\lvert E_{\underline{h}_{obs}} \rvert^{p)}$
0 6 0	3.38	-1.48	1.90	1.85
2 3 6 q)	0.32	3.44	3.76	3.65
4 3 2	-0.68	-1.67	-2.35	2.90
1 3 12	-0.03	-1.62	-1.65	2.52
5 3 $\bar{4}$	0.53	1.45	1.98	2.49
3 3 0	-0.01	-2.18	-2.19	2.20
3 3 $\bar{12}$	0.07	-1.14	-1.07	2.07
6 3 $\bar{12}$	-0.08	2.76	2.68	2.05
5 3 $\bar{9}$	-1.28	-0.78	-2.06	2.65
5 3 $\bar{11}$	-2.29	-0.45	-2.74	2.59

b. Some reflections involved in triples of the type (17.1).

Reflection $\underline{h}$	$E_{\underline{h}}^{(1)}$	$E_{\underline{h}}^{(2)}$	$E_{\underline{h}_{calc}}$	$\lvert E_{\underline{h}_{obs}} \rvert^{p)}$
3 2 4 q)	2.86	-0.58	2.28	2.00
5 2 0	-1.78	0.11	-1.67	2.17
4 2 $\bar{4}$	2.09	0.12	2.21	2.25
6 2 $\bar{6}$	1.78	-0.44	1.34	1.52
3 2 2	-2.04	0.40	-1.64	1.51
1 5 $\bar{2}$ q)	0.39	2.60	2.99	3.01
2 5 $\bar{8}$	-0.19	-1.62	-1.81	2.17
3 5 $\bar{6}$	0.09	-1.74	-1.65	1.86
2 6 $\bar{4}$	-2.38	0.34	-2.04	2.42

p) $\lvert E_{\underline{h}_{obs}} \rvert$ as obtained via the K-curve.

q) As an example; according to the sigma-2 relationship:

$S_{\bar{2}3\bar{6}} \cdot S_{324} \approx S_{15\bar{2}}$ (where $S_{\bar{2}3\bar{6}} = -S_{236}$)

Table 17.5 The consistency with respect to the sigma-2 relationship for four different sets of signs, when also reflections with high k-indices are taken into account.

Triples	1. $\underline{h}_b$ correct, $\underline{h}_c$ correct	2. $\underline{h}_b$ incorrect, $\underline{h}_c$ correct	3. $\underline{h}_b$ correct, $\underline{h}_c$ incorrect	4. $\underline{h}_b$ incorrect, $\underline{h}_c$ incorrect
$S_{\underline{h}_a} \cdot S_{\underline{h}_a'} \cdot S_{\underline{h}_a''}$	consistent	consistent	consistent	consistent
$S_{\underline{h}_a} \cdot S_{\underline{h}_a'} \cdot S_{\underline{h}_b}$ [1]	inconsistent	consistent	inconsistent	consistent
$S_{\underline{h}_a} \cdot S_{\underline{h}_b} \cdot S_{\underline{h}_b'}$	consistent	consistent	consistent	consistent
$S_{\underline{h}_a} \cdot S_{\underline{h}_b} \cdot S_{\underline{h}_c}$	consistent	inconsistent	inconsistent	consistent
$S_{\underline{h}_a} \cdot S_{\underline{h}_c} \cdot S_{\underline{h}_c'}$	consistent	consistent	consistent	consistent
$S_{\underline{h}_b} \cdot S_{\underline{h}_b'} \cdot S_{\underline{h}_c}$	consistent	consistent	inconsistent	inconsistent

[1] The majority of these triples is supposed to be incorrect.

more easily, may be without any trouble at all[x]. This is demonstrated in table 17.5. We distinguish there three groups of reflections:

reflections, $\underline{h}_a$, with indices $k \leq 4$

reflections, $\underline{h}_b$, with indices $5 \leq k \leq 8$

and reflections, $\underline{h}_c$, with indices $9 \leq k \leq 12$.

(We used reflections $\underline{h}_a$, $\underline{h}_b$ and a very few reflections $\underline{h}_c$ in the crystal structure determination). Suppose that the majority of triples: $S_{\underline{h}_a} \cdot S_{\underline{h}_a'} \cdot S_{\underline{h}_b}$ are incorrect and that the majority of the remaining triples are correct. Furthermore, we distinguish four different solutions: all reflections $\underline{h}_a$ are supposed to be correct, but reflections $\underline{h}_b$ and $\underline{h}_c$

[x] Note that a Patterson synthesis will then have a better resolution.

are supposed to be all correct, or all incorrect in turn. Each entry of the table contains a 'consistent' or an 'inconsistent' to indicate whether the majority of signs satisfy or do not satisfy the sigma-2 relationship, respectively. It is clear from table 17.5 that if we restrict the sign determination to reflections $\underline{h}_a$ and $\underline{h}_b$, the most consistent set is set 2, which is incorrect. When reflections $\underline{h}_c$ are also considered, set 2 is no longer more consistent than the other sets. In fact, sets 1 (the correct set of signs), 2 and 4 are about equally consistent with respect to the sigma-2 relationship.

Conclusion: the collection of more intensity data facilitates the structure determination.

CHAPTER 18

In this chapter we will reproduce a paper that has been submitted for publication in Acta Crystallographica.

APPLICATION OF DIRECT METHODS TO DIFFERENCE FOURIER COEFFICIENTS FOR THE SOLUTION OF PARTIALLY KNOWN STRUCTURES

by

Paul T. Beurskens, Th.E.M. van den Hark and Gezina Beurskens

Crystallography Laboratory, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands

Abstract

Partially known structures, which may be a fragment of a molecule found by direct methods, can be solved completely by applying a weighted tangent refinement procedure to difference structure factors, similar to our procedure for heavy atom structures. Often the molecular fragment is found to be misplaced with respect to the symmetry elements: then, the space group symmetry is reduced to P1, and the procedure is used to locate the symmetry elements.

Introduction

In previous papers (Gould et al., 1975, and Van den Hark et al., 1976a) we have described a procedure (for centric, and noncentric structures, respectively) for the solution of heavy atom structures, where the positions of some heavy atoms are known. This procedure is called DIRDIF. Naturally, the procedure can also be used for equal atom structures when a molecular fragment is known.

Application of direct methods for the solution of not too small structures often leads to an electron density map (or E-map) from which a molecular fragment can be recognized. Patterson search techniques or

tangent-formula recycling techniques may then be used to complete the solution of the structure. Often the molecular fragment is in the correct orientation, but shifted with respect to the symmetry elements. The solution of the structure can then be sought by translation functions, or by reducing the space group symmetry to P1, where any position of the molecular fragment is correct by definition.

For these cases the DIRDIF procedure is a very convenient tool, which is illustrated by the following examples.

Examples.

Heptahelicene, $C_{30}H_{18}$, crystallizes in two modifications:

(I) : spacegroup $P2_1$; Z = 4; two independent molecules per unit cell (Beurskens et al., 1976). We had troubles in solving this structure, as could be expected. We managed to find one heptahelicene molecule, incorrectly placed, and we could solve the structure as described below.

(II) : spacegroup $P2_1/c$; Z = 4; one molecule per asymmetric unit (Van den Hark et al., 1976b). We had troubles in solving this structure, partly because of the relatively small number of reflections that could be measured. One heptahelicene molecule was found, also incorrectly placed, and this was used as described below.

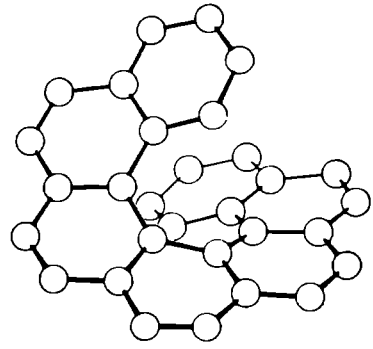


Figure 1. Carbon skeleton of the heptahelicene molecule.

Description of the procedure

The symmetry is supposed to be P1, and the monoclinic reflection data set is expanded using $|F_{obs}(h\bar{k}l)| = |F_{obs}(hkl)|$. One molecule, found by application of direct methods, was used for the calculation of partial structure factors (F_H), and the computer program DIRDIF.D (Van den Hark et al., 1976a) was applied to the difference structure factors to find the most probable values of phases and amplitudes of the structure factors for the remaining three molecules. In contrast to the heavy atom procedure, the overall temperature

factors of the known molecule (B_H) and the remaining three molecules (B_L) were set equal.

The electron density map showed all four molecules, that is: the input-molecule as well as three new molecules. For structure I the two-fold screw axis was easily found, the structure was shifted to bring the two-fold screw axis to its proper position, and the coordinates of symmetry dependent atoms were averaged. For structure II the centre of symmetry was easily found, and shifted to the origin, and symmetry dependent atoms were averaged. The structures then refined rapidly in the proper space group.

Some observations

The input-molecule was also found on the electron density map; this is caused by small errors in scale factor and errors in atomic positions. The 'peaks' in a conventional difference Fourier map will be enhanced by the application of direct methods. From some trial runs it was learned that if more and more atoms are known, these atoms will disappear, and the remaining atoms will have higher peaks on the Fourier synthesis.

Table 1. Distribution of peaks of the four heptahelicene molecules in the Fourier synthesis. (For each molecule the number of peaks that correspond with atoms is tabulated).

		Top 50 peaks	Top 100 peaks	Top 150 peaks
Input molecule	A	14	20	26
Unknown molecules	B	19	28	30
	C	14	26	29
	D	3	16	26
Missing atoms		70	30	9
Spurious peaks		0	10	39

For structure II some numerical results are given in tables 1 and 2. The input heptahelicene molecule (A) is found again equally well as two symmetry related molecules B and C, while the third molecule D had lower peaks on the average.

After shifting the molecules, such as to bring the centre of symmetry to the origin, and after refinement of the structure, it was found that

several input atoms had rather large deviations; the output peaks for the same atoms are much better (see table 2). The deviation for the averaged positions (from all four molecules) are also shown in this table.

Table 2. Comparison of errors in the atomic positions of the heptahelicene molecule before and after the application of the procedure. The tabulated deviations are calculated with respect to the refined atomic positions. The deviations are given for 7 atoms with the largest input error; the numbering of these atoms is as given in the structure report (Van den Hark et al., 1976b).

Atoms	Input molecule	Output A	Averaged positions of A, B, C, D[x]
C(29)	0.78Å	0.11Å	0.11Å
C(18)	0.65	0.06	0.04
C(11)	0.40	0.08	0.04
C(15)	0.36	0.13	0.08
C(16)	0.35	0.04	0.10
C(14)	0.28	0.29	0.09
C(4)	0.27	0.12	0.07
Averaged	0.21	0.15	0.10

[x] see text

There is an important difference between the tangent recycling procedure of Karle (1968) and the present procedure. In the former procedure the same reflections and sigma-2 interactions (used in the initial phasing procedure that resulted in the recognition of a structural fragment) are used again for the refinement of phases. In contrast, quite different reflections and sigma-2 interactions are used in our procedure. Weak reflections, for example, may have a large contribution from the known structural fragment and may be important in our procedure.

Acknowledgement

One of us (Th.E.M. van den Hark) acknowledges support of the Dutch Foundation for Pure Research, ZWO/FOMRE.

References.

Beurskens, Paul T., Beurskens, Gezina and Van den Hark, Th.E.M. (1976). Submitted for publication in Cryst. Struct. Comm.

Gould, R.O., Van den Hark, Th.E.M. and Beurskens, Paul T. (1975). Acta Cryst. A31, 813.

Karle, J. (1968). Acta Cryst. B24, 182.

Van den Hark, Th.E.M., Prick, Peter and Beurskens, Paul T. (1976a). Acta Cryst. A32,

Van den Hark, Th.E.M. and Beurskens, Paul T. (1976b). Submitted for publication in Cryst. Struct. Comm.

When the phases and amplitudes of a large number of structure factors are known, the electron density at each point of the unit cell can be computed. The amplitudes can be obtained experimentally, but not the phases. The so-called phase problem in X-ray crystallography has still not been solved in its generality. However, several techniques have been developed which in many cases lead to the correct solution of the phase problem. In this thesis the so-called direct methods are discussed, which try to solve the phase problem by mainly statistical means.

In part A of this thesis, the use of direct methods for ab-initio phase determination in equal atom structures is described. The emphasis will be on solving crystal structures which can hardly be solved by existing routine direct methods procedures. The difficulties encountered in solving such structures are frequently caused by the use of incorrect sigma-2 interactions in the initial stage of the phase determination process. The procedure presented in chapter 7 makes use of invariant calculations (chapter 6) to select those sigma-2 interactions that should be used or avoided in the phase determination by a sigma-2 procedure (chapter 3). As the accuracy of the invariant calculations is limited, the phase determination is performed by the phase correlation procedure (chapters 4 and 5). When necessary, use is also made of information from the pair relationship (chapter 10). In two crystal structure determinations, in which the procedure was used, problems were encountered. For these structures, the problems could be related to the 'regular' geometry of the molecules in combination with the orientation and location of the molecules in the crystal structure (chapters 11 and 17).

Part B of this thesis presents a description of the use of direct methods to solve structures which are already partially known. The phase information from the known part of the structure is used to find, to verify or to modify the phases of the difference structure factors by direct methods. The result of the application of direct methods is that the remaining (unknown) part of the structure is found much easier and faster. Three procedures are described in part B (DIRDIF.A, DIRDIF.B and DIRDIF.D; chapters 12, 14 and 16 respectively). Originally, the procedures have been developed for the solution of partially known heavy atom

structures, but they also proved to be very useful for the solution of partially known equal atom structures (chapter 18).

In connection with the investigations on direct methods, crystal structures of seven compounds were solved, which are also of interest for chemical investigations at the Chemistry Department of this university (chapters 5, 8, 9, 11, 13, 15 and 17).

Wanneer de fasen en amplitudes van een groot aantal struktuurfaktoren bekend zijn, kan de elektronendichtheid in elk punt van de eenheidscel berekend worden. In tegenstelling tot de amplitudes zijn de fasen experimenteel niet toegankelijk. Het zogenaamde faseprobleem in de Röntgendiffractie is in zijn algemeenheid nog niet opgelost. Wel zijn in de loop der tijd een aantal technieken ontwikkeld waarmee in vele gevallen de juiste oplossing van het faseprobleem verkregen kan worden. In dit proefschrift wordt aandacht besteed aan de zogenaamde direkte methoden, waarmee het faseprobleem statistisch benaderd wordt.

In deel A wordt het gebruik van direkte methoden voor ab-initio fasebepaling in gelijk-atoom strukturen besproken. Het accent ligt daarbij op het oplossen van strukturen die met bestaande routine direkte methoden procedures aanleiding geven tot moeilijkheden. Veelal blijken deze moeilijkheden veroorzaakt te worden door het gebruik van foutieve sigma-2 interakties in het beginstadium van de fasebepaling. In hoofdstuk 7 wordt een procedure beschreven om foutieve fase indicaties in het beginstadium van de fasebepaling zoveel mogelijk te vermijden. Daartoe wordt gebruik gemaakt van invariant berekeningen (hoofdstuk 6) om dié sigma-2 interakties te selecteren die gebruikt dan wel vermeden moeten worden in de fasebepaling met een sigma-2 procedure (hoofdstuk 3). Daar de nauwkeurigheid van de invariant berekeningen beperkt is, wordt in de procedure de fasebepaling uitgevoerd met de fase correlatie procedure (hoofdstukken 4 en 5). Indien nodig wordt ook informatie van de pair relatie gebruikt (hoofdstuk 10). Twee strukturen, waarop de procedure werd toegepast, konden pas na veel moeite worden opgelost. De moeilijkheden konden voor deze strukturen gerelateerd worden aan de 'regelmatige' opbouw van de molekulen in combinatie met plaatsing en orientatie van de molekulen in de kristalstruktuur (hoofdstukken 11 en 17).

Deel B van dit proefschrift geeft een beschrijving van het gebruik van direkte methoden voor het oplossen van strukturen die reeds partieel bekend zijn. De fase informatie uit het bekende gedeelte van de struktuur wordt gebruikt om met direkte methoden de fasen van de verschil struktuurfaktoren te verifieren, te bepalen ofwel te modificeren. Hierdoor wordt het oplossen van het resterende (onbekende) gedeelte van de struktuur aanzienlijk vereenvoudigd c.q. versneld. Achtereenvolgens worden in deel

B een drietal procedures beschreven (DIRDIF.A, DIRDIF.B en DIRDIF.D; hoofdstukken 12, 14 en 16 resp.). Oorspronkelijk zijn deze procedures ontwikkeld voor het oplossen van zwaar-atoom strukturen. Later bleken deze procedures ook zeer bruikbaar te zijn voor het oplossen van partieel bekende gelijk-atoom strukturen (hoofdstuk 18).

Samenhangend met dit direkte methoden onderzoek, werden ook een zevental kristalstrukturen opgelost die mede van belang zijn voor de chemische onderzoekingen die in de sectie Scheikunde van deze universiteit worden verricht (hoofdstukken 5, 8, 9, 11, 13, 15 en 17).

CURRICULUM VITAE

De auteur van dit proefschrift werd geboren op 3 mei 1950 te Sint Hubert. Nadat in 1967 het diploma HBS-B was behaald aan het Elzendaalcollege te Boxmeer, werd in datzelfde jaar begonnen met de studie scheikunde aan de Katholieke Universiteit te Nijmegen. Het kandidaatsexamen (S1) werd in maart 1970 cum laude afgelegd. In oktober van datzelfde jaar werd hem de Unilever Chemie prijs toegekend. De doktoraalstudie omvatte als hoofdrichting Kristallografie en als bijvakken Theoretische Chemie en Capita uit de Wiskunde. Op 5 juni 1972 werd het doktoraalexamen behaald. In die periode heeft hij voor de afdeling Biofysische Chemie ook werkcolleges en practica voor chemie- en biologiestudenten verzorgd. Sedert juli 1972 is hij als FOMRE medewerker verbonden aan het laboratorium voor Kristallografie, waar het hier beschreven promotieonderzoek werd verricht. In het kader van de service verlenende aktiviteiten van het laboratorium voor Kristallografie heeft hij ook meegewerkt aan het oplossen van enkele kristalstrukturen (onderstaande referenties).

Noordik, J.H., van den Hark, Th.E.M., Mooij, J.J. en Klaassen, A.A.K. (1974). Acta Cryst. B30, 833.

van den Hark, Th.E.M. en Noordik, J.H. (1973). Cryst. Struct. Comm. 2, 643.

van den Hark, Th.E.M., Noordik, J.H. en Beurskens, Paul T. (1974). Cryst. Struct. Comm. 3, 443.

Wijnhoven, J.G., van den Hark, Th.E.M. en Beurskens, P.T. (1972). J. Cryst. Mol. Struct. 2, 189.

STELLINGEN

I

Dat Vaska e a aan F^- en aan BH_4^- als liganden een even grote electrone gativiteit toeschrijven, komt doordat zij ten onrechte voor gelijke gevolgen gelijke oorzaken postuleren

L Vaska, W V Miller en B R Flynn, Chemical Communications, 1615, (1971)

II

Stelling 1 bij het proefschrift van Van der Meer "Tegen de door Beurskens en Noordik voorgestelde methode voor het met de symbolische additie oplossen van kristalstrukturen, die zware atomen op algemene positie bevatten, zijn bezwaren aan te voeren" is voorbarig en onjuist

H van der Meer, Proefschrift, Universiteit van Amsterdam, 1971
P T Beurskens en J H Noordik, Acta Cryst A27, 187, (1971)

III

Het valt zeer te betwijfelen dat diododithiocarbamato Fe(III) complexen gesynthetiseerd kunnen worden ook al wordt zulks door Preti en Tosi beweerd

C Preti en G Tosi, Z anorg allg Chem 418, 188, (1975)

IV

De beschrijving van de mogelijke conformaties van een vijfring volgens Altona en Sundaralingam verdient door zijn inzichtelijkheid de voorkeur boven de beschrijving volgens Cremer en Pople

C Altona en M Sundaralingam, J Am Chem Soc 94, 8205, (1972)
D Cremer en J A Pople, J Am Chem Soc 97, 1354, (1975)

V

Hazony en Herber verwaarlozen de bijdrage van intramoleculaire vibraties tot de totale vibratie-uitwijking ten opzichte van de bijdrage van intermoleculaire vibraties op grond van argumenten waarmee eerder het tegenovergestelde aangetoond kan worden

Y Hazony en R H Herber, J de Psysique, Colloque C6, 35, C6 131, (1974)

VI

De constatering van Levitzki e a dat oxidatie van Cu(II) peptide complexen leidt tot specifieke breuk in de peptide keten tussen het C en N atoom van het derde aminozuur, geteld vanaf de eindstandige NH_2-groep van het peptide, ondersteunt de aanname van Bour e a dat de lage oxidatie potentiaal van biuretato Cu(II) complexen te wijten is aan een hoge electronen dichtheid op het centrale koper atoom

A Levitzki, M Anbar en A Berger, Biochemistry 6, 3757, (1967)
J J Bour, P J M W L Birker en J J Steggerda, Inorg Chem 10, 1202, (1971)

VII

De veronderstelling van Ollie B Bommel, geuit terwijl hij als proefpersoon optrad bij een experiment van professor Sickbock met aangrenzende tijdruimtes, dat hij daadwerkelijk meegewerkt zou kunnen hebben aan zijn terugkeer in het heden, is onjuist

Marten Toonder, De Bommellegende, dagstrips 4047-4112

VIII

Bij de uitgave van een postzegel ter stimulering van het sparen, verdient het aanbeveling op de afbeelding elke associatie met de posttarieven te vermijden

Nijmegen, 17 juni 1976 Th E M van den Hark